505

R 815

# DIRECTORATE OF DISTANCE EDUCATION

ACL. MKU
01165

LABOUR MANAGEMENT

M.L.M., SECOND YEAR

# SOCIAL RESEARCH AND STATISTICS

MADURAI KAMARAJ UNIVERSITY
PALKALAI NAGAR, MADURAI - 625 021.

All copyright privileges are reserved

Printed by : **M.M. 3114, Madurai Kamaraj University Co-operative Printing Press Ltd.,**
Palkalainagar, Madurai - 625 021. Phone : 2458679 Copies : 1000 10/2007

Master of Labour Management
Second Year

# Social Research and Statistics

### Syllabus

## Part A : Social Research

Scope and meaning of social research - Social Surveys, Planning and execution - stages in research project - methods of data collection, documentation and bibliography, schedule, questionnaire and interview - pilot study - processing of data, classification, editing, tabulation, report writing etc.

## Part B : Statistics

Method of presentation of statistical information - sampling and its distribution - Testing of hypothesis - statistical treatment of wage problems, consumer price index, employment, production and consumption - National income and social accounting - Types of sampling and techniques of compilation - Government publications relating to agriculture, trade, industry, labour and census.

## Books for Reference

| | |
|---|---|
| Goode and Hatt | : Methods in Social Research |
| Wilkinson and Bhandarkar | : Methodology and Techniques of Social Research |
| Pauline, Young | : Scientific Social Surveys |
| Parsons | : Thesis and Project work |
| Uuie and Kendall | : An Introduction to the Theory of Statistics |
| I LO | : International Standardisation of Labour Statistics |
| I LO | : World Labour Report |
| I LO | : Year Book of Labour Statistics |
| Golden | : Statistical Methods |

# Master of Labour Management

## Social Research and Statistics

## SCHEME OF LESSONS

### Lesson - 1 Meaning and Scope of Social Research

What is Research? Motivating Factors of Social Research - Basic postulates of Social Research - Qualities of social researchers - Methods of Social Research - Value of Social Research in India.

### Lesson - 2 Social Surveys

Definition - Difference between Social Survey and Social Research - Object of a social survey - Types of surveys - planning and execution of a social survey - Merits and limitations of survey method.

### Lesson - 3 Stages in a Research Project : Data Collection

Primary Data - Secondary Data - Sources of Primary data - Secondary sources - Documents - utility of secondary sources

### Lesson - 4 Questionnaire and Schedule

Definition - purpose - Form - Characteristics of a Good Questionnaire - Model Questionnaires - Disadvantages and importance of Questionnaire method.

### Schedule

Meaning - features - objectives - Types Construction of a schedule - merits and demerits of schedules - Distinction between schedule and questionnaire.

### Lesson - 5 Interview Method

Meaning, objects and types of interview - Technique (Process) of interview - merits and demerits of interview method - Guidelines for successful interviewing.

### Lesson - 6 Editing, Classification and Tabulation

Editing and coding of the data - classification of data - Quantitative classification - Tabulation of data - types and uses of tables.

### Lesson - 7 Report Writing

Planning and organisation of a report - Formation of a satisfactory outline - conclusions and generalisation - The Three drafts - Bibliography.

### Documentation and Bibliography

Documentation : Meaning and components - Footnotes, Bibliography - quotations and translations, tables, charts and diagrams and appendices - prefaces.

## Lesson - 8 Methods of Presentation of Statistical Information - I : Diagrams

Usefulness and limitations of diagrams and graphs - General rules to be followed in drawing diagrams and graphs - Bar Diagrams - Area Diagrams - Pie diagrams - Pictograms

## Lesson - 9 Methods of Presentation of Statistical information - II Graphs

Histograms - Frequency Polygon - Frequency Curve - Ogives

## Lesson - 10 Sampling and its distribution

Meaning of population and sample - Census versus sample method - Sampling Distribution - Standard Error of Mean

## Lesson - 11 Testing of Hypothesis

Procedure of testing a hypothesis - Sampling of Attributes - Sampling of variables (large and small samples) - Chi square test

## Lesson - 12 Statistical Treatment of Data - I

Wages - Consumer price index numbers - employment - production

## Lesson - 13 Statistical Treatment of Data II - National Income and Social Accounting

National Income : Definition, methods of estimation - concepts of national income - Estimation of National Income in India - Difficulties in its estimation - Recent improvements in estimation - Social Accounts - uses of Social Accounting Framework.

## Lesson - 14 Indian Official Statistics

Nature and structure of the official statistical organisation in India - Agricultural Statistics - Trade Statistics - Industrial Statistics

## Lesson - 15 Indian Official Statistics

Labour Statistics - Population statistics

*Lessons Compiled by :*

**Dr. Jagadeesan,**

Department of Labour Studies,

DDE,

Madurai Kamaraj University,

Madurai - 625 021.

## LESSON - 1

# MEANING AND SCOPE OF SOCIAL RESEARCH

## What is Research?

The Webster's International Dictionary proposes a very inclusive definition of research as "a careful critical inquiry or examination in seeking facts or principles, diligent investigation in order to ascertain something." This definition, is not specific and precise enough to give us a more complete and a more clear idea of what research involves. Improving upon this definition, it may be stated that research refers to "a critical and exhaustive investigation or experimentation having as its aim the revision of accepted conclusions in the light of newly-discovered facts." The researcher is constantly concerned with re-searching the accepted conclusions of his field, ie. the theories with differing levels of generality and degree of confirmation (trust worthiness) obtaining at a point in time. He does this researching for facts of the empirical world that confirm or falsify one or more predictions generated by his accepted conclusions. Thus, researching may turn to the construction of new theories to take the place of those no longer able to fit the data of the empirical world. Research stated otherwise, is a systematic attempt to push-back the bounds of comprehension and seek beyond the horizons for our Knowledge same 'truth', same reality shrouded in a subtle way and consequently, to keep on extending and also consolidating these horizons without end.

We shall attempt a clause analysis for the sake of a comprehensive understanding of the term 'research'.

### (a) Manipulation of things, Concept or Symbols

Science deals with things. In a laboratory, the physicist deals with things. In order to know how things respond under specific conditions, what the common denominators that are discernible in their responses under varied conditions and what law they abide by etc. The scientist subjects them to purposeful control e.g. starving the rats for a specific period or keeping the balls on different degrees of incline, etc. This purposeful handling we call manipulation which is an aspect of experimentation.

The scientist is not dealing with things alone but concepts or symbols. For the things that a particular science deals with, it has and must have some terms designating the things about which a science tries to make sense are its concepts. If the scientist is dealing with things only he would be working and constrained to remain at the level of concrete and his results would be limited to the particular thing subjected to study and none else. The scientist dealing with things is actually dealing with concepts which symbolize them and their various properties. The physicist in our example is thus dealing with the abstract notions of spheres, circles, velocity, acceleration angles, pluses and minuses. The psychologist is dealing with abstract notions of organism, hunger, deprivation and problem solving potential etc. All these abstract notions relate to things that are being dealt with but are not things themselves. These are the mental short-hands for things. Some concepts or symbols are abstract forms denuded of content. It is as though the scientist working with them was working with certain classes of things at a higher level of abstraction. Concepts have thus the relevance of things. Research involves manipulation of things and concepts at the hands of the scientist.

## (b) For the purpose of Generalizing

The manipulation or purposeful control of things, concepts or symbols are engaged in with a definite purpose which is to arrive at statements of generality. Of course, generality is a matter of degree and the conclusion thus arrived at on the observations under exercise of manipulation will vary in their degree of generality. The level of generality may be higher or lower the significance or worth of a research may perhaps be judged in such terms but an absence of generality hardly characterizes research.

## (c) To Extend, Correct or Verify Knowledge

Generalization has in turn certain effects on the established corpus of knowledge. It may extend the bounds of knowledge existing at a point in time. It may, if it does not fit the existing knowledge, exert pressure to bring about amendment in order to correct it. If the generalization fits the established groups of knowledge, it lends added confirmation to it.

The role of general propositions in enlarging or extending systematic knowledge may take various forms. The generalizations can be instructively used to understand certain phenomena or aspects of them that were not examined before. A new observation incorporated in the generalization may afford a basis for understanding a phenomenon better. The phenomenon may reveal itself as non governed by the same laws governing certain other phenomenon. This means that the phenomenon in question can be included in a class of phenomena which may be a unit of a still larger class of phenomena. In this manner a new general conception is brought within the grasp of the established corpus of knowledge in a meaningful way. Consequently, the bounds of knowledge is enlarged not just quantitatively (ie. more units covered under a conceptual scheme) but also qualitatively (ie. in depth of understanding and in bettering predictions). For example, a new idea such as "relative deprivation" may serve as a lever for examining the relation between rise in wages and productive efficiency of a worker. The problematic observation that a wage-rise does not always accompany productive efficiency (it may at times even hamper it) can be properly understood by means of the concept. This behaviour of workers then appears to be a manifestation of reference group behaviour. What we have gained is a deeper understanding of a particular phenomena.

Another way in which a general conception may enlarge systematic knowledge is by bringing to light the inconsistencies in the existing corpus of knowledge and attempting to reconcile these. For example, the findings of studies of "authoritarian" and democratic class-room atmospheres on learning are not consistent. A new general conception helps reconciliation by pointing that whatever be the class-room atmosphere, the important thing is the extent to which the teacher behaves according to the expectations of the students.

Another way in which a new general conception extends knowledge is by pointing out the gaps in the existing corpus of knowledge and attempting to bridge these up. These gaps have a reference to the failure or inability on the part of a conceptional scheme or theory to explain or account for certain aspects of a phenomenon. This inadequacy of the theory represents a gap in knowledge. The role of a new general conception may be to bridge this gap and in as much as this is successfully done, knowledge gets extended.

The general conception or finding may have another impact on the established body of knowledge, i.e., of correcting the errors in it. If the general proposition derived through manipulation of things, concepts or symbols is discordant with what one would expect on logical grounds, on the basis of the existing body of knowledge or thereby the only alternative (since the empirical observations and corpus of knowledge cannot stay divorced from each other in mutual indifference) is to effect corrections or modifications in the original corpus in the light of new evidences in case, the existing system of knowledge has no place for the new observations.

Research involves testing whether the empirical observations presented as general statements are in accord with the predictions made on the basis of the prevailing system of knowledge. If the observations are in accord with these predictions, the corpus of knowledge acquires greater strength and confirmation. An important aspect of scientific activity is verification of conclusions which have found a place in the established system of knowledge. In science, the established facts must be really established for they are the rallying points, the basis for the superstructure.

### (d) Knowledge may be used for construction of theory or practice of an art

The extended, corrected and verified knowledge has two possible uses to which persons may put it.

(a) Theoretical

(b) Practical

Knowledge thus acquired may be used for building up theories. Knowledge may be organised into propositions. It may then be meaningfully articulated to form a more abstract conceptual system affording estimations about a class of things or phenomena under specified classes of conditions. Such a use of knowledge is known as theory oriented and the activities of a scientist who seeks knowledge for the sake of building. Theories of non-utilitarian import are often known as 'Pure, basic or 'theoretical' research. Knowledge for the sake of knowledge, i.e. for the satisfaction of 'knowing', characterises such as attitude.

Knowledge on the other hand, may be used as a means to some practical or utilitarian goal not just as an end in itself. For example, the knowledge about people in respect of certain things may be used to work towards their welfare, ameliorate their practical problems or exploit them for certain personal or ideological ends. The research which seeks knowledge for using it to serve some practical ends are often called applied or 'action-oriented or practice-oriented.

## Research Process

The research process consists of a number of related steps which are essentially interdependent. Research starts from the conception of a problem. Then collection of facts and information will follow and these will be processed by scientific techniques and tools to arrive at a reasonable conclusion. The following steps are involved in research.

(i) Formulation of the problem in tune with the purpose of the study.

(ii) Description of the research design.

(iii) The methods of collection of data.

(iv) Results of the study.

(v) Policy implications and the conclusions.

The formulation of the research topic or project has to be based on a number of considerations which are as follows :

1. The problem which requires solution must be from within the area of interest and feasibility of the research worker. Topics requiring immediate solution or involving controversy may be given preference over other ordinary topics.

2. The entire research project may be further broken up into a number of small and distinct projects of manageable size.

3. It is essential to formulate the hypothesis which will be helpful for the collection and analysis of data. The hypothesis has to be subsequently tested by facts.

4. The definitions of the concepts and symbols involved must be given precisely. The definitions must be operational and the conceptualisation must be simple but clear without any ambiguity or contradiction.

5. The methods and analytical tools should be specified, applied to find out the result of the study.

It is necessary to relate the generalisation to the existing body of knowledge. This involves a thorough knowledge of existing literature and a foresight of the emerging trend.

## Social Research

A social research encompasses scientific investigations conducted in the field of social sciences and also in behavioural sciences. Social research is a very broad category within which there are many sub-classes. According to P.V. Young, a social research is "the systematic method of discovering new facts or verifying old facts, their sequences inter-relationships, casual explanations and the natural laws which govern them". Social research involves the application of scientific method for understanding, studying and analysing social life in order to modify, correct or verify the existing knowledge as a system.

Social research has some distinct characteristics. It studies human behaviour and social problems. The main idea behind such research is to discover new interrelations, new knowledge, new facts and also to verify old ones. In a progressive science, it is essential to verify constantly the old facts and conclusions. Such verification is warranted, if the technique of analysis improves the concept itself, changes or new facts or data become available. Social research tries to find out the causal connection between various human activities and the natural laws governing them by

means of logical and systematic methods. Human behaviour may be motivated by certain rules and laws. The main purpose of social research is to discover those laws which can be the proper guidelines for studying human conduct and behaviour.

## Objectives of Social Research

1. Social research aims at understanding the human behaviour and its interaction with the social institution.

2. The purpose is the acquisition of new knowledge in order to gain insights into existing problems. The collection of new facts is also essential for the progress of science.

3. P.V. Young observes that the primary aim of social research is to understand social life and thereby to gain a greater measure of control over social behaviour. It is necessary to investigate into the causes of certain social phenomena in order to cure the social events.

Another purpose of social research is to try to improve our tools of analysis or to test these against complex human behaviour and institutions. Social research should, in the short run, try to develop special theories with limited applicability; but in the long run, attempts may be made to put forward more general theories.

## Motivating Factors of Social Research

P.V. Young has mentioned four motivating factors of social research.

### 1. Curiosity about unknown

"Curiosity is an intrinsic trait of human mind and a compelling drive in the exploration of man's surroundings". It is a natural instinct in mankind. Even a small child is curious about the unknown objects that it notices around it and tries to understand them in its own ways. The same curiosity drives scientists to explore the unknown factors working behind the social phenomena.

### 2. Desire to understand the cause and effect of widespread social problems

The search for cause and effect relationship has been more relentless than almost any other scientific effort upon which human energies have been spent. More and more research is undertaken to dispel doubts and uncertainties which result from inadequate conceptions of underlying factors shaping social processes.

### 3. Appearance of novel and unanticipated situations

Man is often faced with many acute and difficult social problems. A social scientist sets down dispassionately to find out their real cause and thus evolves a lasting solution to such intricate problems.

### 4. Desire to discover new and test old scientific procedure as an efficient way to gain useful and fundamental knowledge

Of late there has been a growing emphasis upon the use of quantitative or statistical methods in social research in order to make it more definite and mathematically precise.

## Types of Social Research

There are mainly three types of social research,

(i) Fundamental research (ii) Applied research (iii) Quasi-social research.

### 1. Fundamental Research

These researches deal with the fundamental principles of sociology. They may be conducted either for the verification of some old theory or establishment of a new one. Need for such a research arises with the emergence of new problems and new situations in the society and the questions that the old theories fail to answer satisfactorily. The dynamic character of social phenomena require that old theories should constantly be modified in order to make them more perfect. Emergence of new theories is also made possible by the improvement in the methods of research and its tools of analysis. With more accurate and reliable tools of research, discovery of new fundamental facts regarding organisation and working of social institutions is possible.

### 2. Applied Research

Applied research deals with the possibilities of application of the results of fundamental research to social problems. It, therefore, deals with social therapy or social engineering. Applied research generally takes the form of social surveys. The major difference between an applied and fundamental research is that whereas the latter deals with the fundamental principles guiding the working of human society, the former pertains to more immediate problems. Fundamental principles have a much wider application while the scope of applied research is narrower. Fundamental research provides the basis upon which the whole superstructure of applied research is built up.

### 3. Quasi-Social Research

Social sciences have many common boundaries among them. Naturally there are a large number of problems, which we may call as border line problems. An applied research in all these problems may be termed as quasi-social research. Socio-economic, socio-psychological and socio-anthropological researches come under this category. The distinction between a pure, applied social research and quasi-social research is thus dependent upon subject matter and not on the use of tools or fundamental principles.

## Basic Postulates of Social Research

The application of scientific methods of sociological research is based upon certain assumption which may be classified as below.

### 1. Existence of cause and effect relationship

It has to be accepted as a basis of social research that their cause and effect relationship between various social activities. For example, social crimes may be caused by poverty and frustration. In other words, when there is poverty and frustration it will result in social crimes. Therefore, poverty and frustration have to be eliminated first, in order to eliminate crimes.

## 2. Existence of Sequence or Law in Social Activities

Another assumption is that various social activities do not occur in a haphazard random manner. There is some system or some trend behind them. If this trend or system is located it is possible to predict the future course of social phenomena. For example, it is proved that population of a country increases by some specific rate, then the population of the place in any future year can safely be predicted.

## 3. Possibility of Detached Study

Although man is part of the society which he is investigating, yet it is possible for him to study it as apart from him. His own feelings and emotions would not be reflected in the study, although it is a very difficult task, it is not impossible.

## 4. Existence of ideal types

In society every one is not entirely different from each other. People may be grouped into fairly homogeneous classes known as 'ideal type'. The deductions drawn from the study of the group, may be made applicable to the whole type, whenever it may exist.

## 5. Possibility of a Representative Sample

The assumption is similar to the previous one. It is generally assumed that a sample representative of the group may be drawn and the deduction from the study of the sample may be made applicable to the whole group. The social researcher has to rely upon the sampling technique under the assumption that a sample may be representative of the whole group.

## Utility of Social Research

Social research is a curious course of enquiry, satisfying human desire to know the facts hitherto unknown or obscure. It increases the social stocks of knowledge and so also its flow in various branches of social dynamics. It focuses our attentions to social reality and social events. It strengthens our desire for truth and opens up before our eyes, hidden social mysteries. Social research helps us to scientifically judge and accept a phenomenon by discarding all superstitions, cliches, orthodox beliefs and ignorance covering the phenomenon.

## Utility of Social Research.

## 1. Control Social Phenomena

Knowledge is power. Social research can equip one with detailed and sufficient knowledge about the working and organisation of a society and its institutions. In this way, it can give greater power of control over the social phenomena. The growth of leadership is facilitated by social research.

## 2. Social Planning

Any planning would depend on the systematic knowledge of the resources, problems desiderata and objectives of a society. Social research can give us the detailed picture of the contemplated aspects of social planning.

### 3. Social Understanding and Good will

Social research pinpoints the need for interdependence among different social groups. It reveals the truth by dispelling superstitions, ignorance and taboos. It gives a clear weightage to independence and reasoned opinion and helps to promote good will and understanding. It brings out the unity among diversities and helps to strengthen social cohesion.

### 4. Social Growth

Social research points out to society a right and normative way of development by pinpointing the evil effects of the wrong course of action. Given the constraints, the direction for social growth can be given by social research by study of societal organisations, institutions, values, motivation and so on. The fact that social research can lead to better social control, planning and cohesion by itself goes to prove that it may also lead to better social order, functioning and growth.

### 5. Social Prediction

Social research affords a sound basis for prediction in a large number of cases. Although the prediction cannot be perfect due to a variety of limitations and constraints in the majority of cases, a reasonable result may be ascertained on the basis of the reasonable value of prediction, better social control and planning can be attempted through social research.

### 6. Modernisation of Tools and Techniques

With the progress of social research, the tools and techniques of analysis become modernised and up-to-date. Necessity is the mother of innovation. Due to the exigency of new situations, better tools for social research may be devised and better and more efficient results may be expected.

### 7. Social Welfare

Social research can unfold and identify the causes of social evils. In this sense, it can help to take the necessary action for eradication of social evils. Social research can give sound guidelines for appropriate measures of reforms and social welfare. Social research can indirectly maximise social welfare.

## Qualities of Social Researchers

A researcher is one who knows more and more about less and less. This shows that a good researcher must have a specialised area of interest. It is almost impossible to delineate the actual qualities of a good researcher. However, some broad qualities of a good researcher may be indicated as follows :

### 1. Scientific Mind

A researcher must have a scientific frame of mind. He should not be influenced and guided by pride and prejudice, and by the superficial facts. He must give up personal likes and dislikes. He must be bold enough to discover new things and to discard superstitions and taboos.

## 2. Seeker of Truth and Knowledge

A researcher must be a seeker after truth. The lure of discovering the unknown fact is the starting point of the research.

## 3. Alertness, Insight and Imagination

A scientific mind must always be alert to appreciate minutest changes in situations. A researcher should be accurate in observation, quick in perception and must have precision of statement. His mind must be thoroughly disciplined. He must have a high degree of imaginative power and be able to catch the clues by probing deep into the matter.

## 4. Quick Power of Understanding

A research worker should have the ability to grasp things quickly so that he is in a position to make the best of his research.

## 5. Trained and Educated

A researcher must have intimate knowledge of his area of research. The concepts, symbols and the implications of his project must be very clear to him. He must have sufficient experience and training to understand, analyse and tackle the problem. A researcher must have an analytical mind. Simple description of problems is no research. The results must be found out by rigorous methods. Therefore, he must be acquainted with the sophisticated and latest technique of research. He must be very clear about the methodology he is following.

## 6. Patience and Perseverance

Research is a curious mixture of success and failure. It is an intellectual exercise requiring endurance and patience. He requires supreme courage of conviction.

## 7. Objectivity

A researcher must be objective in his approach. A scientist must try to avoid sentimental and emotional interpretation of his result. He must have an open mind.

## 8. Knowledge of Scope and Limitation

A researcher must clearly know as to what he is going to show and how he can prove his case. He must have the capacity to clearly distinguish. But he must not be too ambitious. He must know his limitations and constraints. A good researcher must be able to answer who, what, when, why, where, how and so on, in relating to his own topic of research. He must know where to begin and where to end.

## Methods of Social Research

It is difficult to give an exclusive classified list of the methods used in social research. Most of the so called methods are only different ways of collecting data. Then again there is a good deal of overlapping the techniques and exclusive classification is rendered all the more difficult. The various methods can however be grouped up under the following heads.

## 1. Qualitative and Quantitative Methods

Research methods may roughly be classified into qualitative and quantitative according to the nature of data used. Qualitative, logical or philosophical method of social research is much older and even now some of the sociologists regard this as the only method for fuller study of social phenomena. Under this method, the collected information is in the form of quantitative description of self experience. The inferences drawn in this method are naturally based upon logical or philosophical rules rather than the facts.

A quantitative or statistical method is used when the phenomena can be expressed in the form of numerical data. The quantitative data is subjected to various kinds of statistical analysis to establish relationship or test the validity of a hypothesis. 'Increasingly more end use is being made of the quantitative methods and attempts have successfully been made to convert descriptive social phenomena into numerical terms. Statistical analysis is mathematical in nature and relies upon such measurements as coefficient of correlation, association of attributes, covariance measure and the nature and extent of relationship between two or more variables. Statistical methods are based upon the analysis of mass data and for valid generalisation, fairly large number of cases is essential.

## 2. Experiment and Field Study Methods

Under this method the subject under study which may be an individual or social group is subjected to certain external stimuli and the effect of the same is watched to find out its influence. The basis of experiment is, thus, capacity to manipulate the variable of causative factors with a view to estimate the casual relationship.

Experiments can be conducted in the setting as well as in the laboratory. The field experiment has been used for studying a variety of techniques, training methods, effects of group decision and participation. Experiments are mainly psychological in nature.

A major part of social phenomena still remaining outside the scope of experimental research. As Lundberg points out, "the social scientist will probably never be able to bring into his behaviour in a test tube under varied conditions".

## 3. Case study approach

Case study approach involves a deeper investigation of a single unit, may be an individual, a family, an institution, a district, a community or any single event selected for intensive examination. It is an approach which views any social unit as a whole. It is a way of organizing social data so as to preserve the unitary character of the social object being studied.

## 4. Evolutionary and Comparative Methods

The methods of scientific study can further be classified into two parts

i) Evolutionary or Genetic method.

ii) Comparative method.

## (i) Evolutionary Method

Under this we try to trace the history of development of a phenomena and thus try to locate the unity. In their present state two facts may seem to belong to different groups because of the different stages of development but a trace of their earlier form may lead to the discovery of unity. This method is very useful for the study of human customs and institutions. The method was first used in Comparative Philology. (Science of Language) in 18th century. The hypothesis in this respect was the unity of mankind and common language. Darwin's origin of species is another example. Infact the whole theory of biological evolution rests on the application of Evolutionary Method.

### The functions of the Evolutionary Method :

(1) It indicates the main steps or stages through which the development has taken place.

(2) It indicates the reason for the various changes constituting the several stages in the suggested line of development.

## (ii) Comparative method

This method aims at discovering the truth by comparative analysis of different groups. It has found more favour in the sociological research. Social groups widely differ from each other and a comparative study of a phenomena in different groups will reveal the basic laws working within it. For example, a study of marriage among different social groups would help to reveal the basic unity and find out the relationship between marriage and social setting of the particular groups. Comparative study removes much of the misconception born of biased thinking and faulty generalisation and makes the theory more broad based.

Generally comparative method is used where the investigator believes that the phenomena in question is not the product of evolution but in most sciences the two methods are used almost simultaneously by different schools of thought in sociology itself. Social Anthropology and Comparative Sociology may be given as illustrations of the historical and comparative methods being used in the same discipline.

### Value of Social Research in India

India is a welfare state. In addition to general social services like education, health, and specialised social services are intended for the benefit of the weaker and vulnerable sections of the population. The social welfare services need professionally competent personnel serving the cause of the handicapped of various kinds. Such services in order to be beneficial and fruitful have to be based on reliable information about people whose welfare is being contemplated. Welfare, in the true sense, implies giving just the right kind and degree of support to the handicapped in view of his problems with a view that the afflicted acquires enough strength, capability and motivation to help himself grow. Such help or support that will not corrupt a person or a class of persons but will enable him can only be administered on the basis of data afforded by scientific social studies. Legislation, which is a means to amelioration, must be based on a reliable assessment of the needs of people and not on mere assumption.

As an up coming industrial nation, India is faced with certain characteristic challenges which need an urgent coming to grips with. The first and the most obvious social alteration of the country under industrialisation was in the productive organisation itself. Compromises with traditional norms of conduct, their effect on workers' efficiency and the extent to which such compromises impede full social transformations are important questions concerning work relation in India. Industrial technology has imposed some minimum requirements on the organisation of production units. The bureaucratization of labour force-an accompaniment of industrialisation-presents various transitional problems not only for factory recruits but also for the elite group. The need for subjective attitudes and objective incentives favorable to novel jobs and conditions of employment is implicit in the requirements of new work relations and forms of productive organisation. Various obstacles to fuller commitments to industrial work are evident. Industrialisation involves gross changes in the occupational structures by the economic sector and finer changes by demand of specialized skills. A vital economic problem of India is the shortage of capital. Thus, budgetary behaviour of people need to be taken into active account.

The population explosion and the rapid rate of urbanization seem to be connected with industrialization in India. Infant mortality has gone down; there is in evidence a high intensity toward migration, age composition has changed and with it has emerged the problem of dependency. All this has serious implications for vital spheres of social life. In India cities seem to have grown at a rate surpassing both expansion of employment opportunities and public services.

Technological modernization in India seems to be linked with profound changes in the family and kinship organisation. A common symptom engendered by industrialization appears to be the disintegration of the family. Industrialization has given rise to many a problem of community organization, e.g. problem of residence, transportation, congestion and slums constituting one environmental source of a number of problems of social control. There are crucial questions of social strategy involved in formulating an educational policy in view of industrial demands. The correlation between education, supply with manpower demands is hardly perfect. The mass media of communication, supply the technical basis for promotion of a national public participation in political events and policy-making. Political participation seems to have been organised around unions or other occupational groups for a variety of interests and loyalties. Industrialization seems to have influenced, in important ways, the religious organization and beliefs.

Industrialization has provided a new set of social positions and new criteria of placement and valuation. It has, therefore, resulted in the complication of the system of stratification. This has been responsible to some extent in status-ambiguity.

The manifold and multi-faceted implications of industrialization for India typically spell out the need for scientific social research on a fairly large scale to provide for a systematic understanding of the nature of problems, a sound diagnosis and an effective treatment plan. Infact, the complexity of the situation (of which little is known) calls for some kind of an inter-disciplinary collaboration among the researchers from specialized disciplines - economists, political scientists, sociologists, psychologists and anthropologists etc. making for an effective multi-pronged attack on these maladies.

## Model questions

1. What are the motivating factors of research?
2. Explain the types of social research.
3. What are the basic postulates of social research?
4. Discuss the utility of social research.

## Suggested Readings


1. Pauline V. Young : Scientific Social Surveys and Research, Prentice Hall of India Private Limited, New Delhi 1977, Chapters 2 and 3.
2. Sadhu Singh : Research Methodology in Social Sciences, Himalaya Publishing House, Bombay 1985, Section 1.
3. S.R. Bajpai - Methods of Social Survey and Research, Kitab Ghar, Kanpur - 3 (Chapter 2)
4. Wilkinson & Bhandarkar - Methodology and Techniques of Social Research, Himalaya Publishing House, Bombay 1979 (Chapter - 1)
5. B.N. Ghosh - Scientific Methods & Social Research, Sterling Publishers Private Limited, New Delhi, 1982 (Chapter - 18)


### LESSON - 2

# RESEARCH DESIGN AND SOCIAL SURVEYS

When the selection of the problem and formulation of hypothesis are completed, the third step of scientific research is to prepare a research design. Research design is a framework or plan for a study that is used as a guide in collecting and analysing the data. It is a blue print that is followed in completing a study. Just as a house built without a blue print is likely to be subjected to midstream alterations incurring more cost in the construction, research conducted without a research design is likely to cost more than the research properly executed using a research design. According to Sellitz and others, research design is the arrangement for collection and analysis of data in a manner that aims to combine relevance to research purpose with economy in procedure. Research design indicates to the entire process of planning and carrying out a research study prior to the actual field works so that the researcher would go ahead/proceed with his investigation, clear ideas of setting and other steps concerning the research work. The quality of a design of research is judged by the degree of accuracy and practicability, attainable on the level of relevant evidences sought.

A poor research design may produce insignificant and erroneous conclusions. So, special care must be taken while developing the research design. Each type of research requires a particular type of design. Type of research is based on the purpose of research and the nature of the problem chosen for research. Therefore, the research design also depends upon the nature and dimensions of the research problem and the purpose and requirements of the whole process of research.

A good research design should fulfil four conditions as given below

### 1. Objectivity

The research design should permit the use of measuring instruments to measure the truth or fact that it was intended to find out. The instrument is fairly objective in which every researcher using it, arrives at precisely the same result. The objectivity of agreement between the final scores of the procedure may be judged by the degree assigned to different individuals by more than one independent observer. In the case of more 'subjective' measurements, the final evaluation and report of different observers will not agree with each other.

### 2. Reliability

Reliability means the 'consistency' throughout a series of measurements. If a respondent gives out a response to a particular item, he is expected to give the same response to that item whenever he is asked subsequently. The investigator should frame his items in such a way that the respondents could give only the genuine response. If the respondents keep changing the response to the same item when asked repeatedly, it will mean that the 'reliability' of the design is less.

### 3. Validity

The instrument should measure what it is intended to measure. For instance, an attitude scale intended to measure the attitude of respondents towards war should measure only, his attitude towards war and not anything else.

### 4. Generalization

Generalization refers to how best the data collected from a sample can be utilised for drawing commonness which may be applicable to a larger group. The conclusion of the study conducted on a small sample should be good enough to be applicable to a larger population, selecting the sample, application of a proper tool and use of appropriate statistical analysis.

Hence, in preparing the blue print, that is, design of the study, the researcher, should ensure all the four above mentioned factors. There are various types of designs depending on the research purpose. The purpose of research may be differentiated into three broad groups depending upon the nature of the problem. They are as follows.

1. Experimental design is the one in which the researcher tests hypothesis of a causal relationship between variables. There are different varieties of designs. They can be brought under two broad categories, viz, experimental design and non-experimental designs.

2. When the research aims to portray accurately the characteristics of an individual or a phenomenon or an event, the design adopted is known as descriptive or exploratory design.

3. In certain other types of research, the investigator wants to gain more familiarity with the phenomena and achieve new insights into it. Hypotheses are likely to be formulated. This type of design is referred to as exploratory or formulative design.

## Experimental Designs

A design programmed for conducting the experimental research is called as experimental design. Different types of experiments can be conducted depending upon what we want to know. Experimental designs refer to plans outlined for experimental research.

The most important types of experimental designs are discussed below.

In experimental design, the researcher asserts that a particular occurrence (X) will determine the occurrence of another event. That is 'X', is the cause of 'Y'. It is also possible that whenever 'X' occurs there is the likelihood of 'Y' to occur. This is called concomitant. There are also the time factors to take note of that is, 'Y' follows 'X' - and not - 'X' follows 'Y' in which 'X' can not be the cause of 'Y'. In the physical world the direct causation is possible like-the boy hit the bottle with a rod and the bottle broke. Here, the hitting with the rod is the cause for the breakage of the bottle but in social events contributory or partial causation is recognized. The cause of an event may be indirect and probable. Establishing the causal relationship between two variables in social sciences is done with the help of experimental design.

Experiments are studies in which the researcher has controlled independent variable over the assignment of subjects to different conditions. The ideal experiment is conducted in a laboratory in which the effects of the independent variables are assessed while other possible effects are ruled out. However, experiments can not be conducted in 'real life' situation if the researcher has not enough control over the environment (In experiments we identity two different variables). One is dependent variable in which the changes are results of the influence of the independent variable or variables. The independent variable is the one that is manipulated or treated in a study to see what effect it will have on those variables proposed as being dependent on it. For example, sex is an independent variable and education a dependent variable.

To conduct an experiment, we should have two groups-a-control group and an experiment group. The control group provides essential protection against the possibility that, factors other than experimental treatment have influenced the dependent variable. These two groups could be formed with subjects either by (a) randomization or (b) matching. Randomization involves random assignment of subjects to these two groups. The assignment procedure might give each subject the same chance of being assigned to any one of the two groups. This could be done just by flipping a coin or giving random numbers. In the case of matching, certain background factors of the subjects with same qualities and features are taken into account. Care is taken to match them perfectly. Matching increases the sensitivity of the experiment. However, matching of groups is a difficult process.

There are numerous types of experimental designs to suit the nature of the study. The two most significant and basic forms are after only design and before-after design.

## a) After only design

Individual subjects are randomized to make two groups. These two groups are not measured with respect to the dependent variable before exposure to the experimental variable. The experimental group is exposed to the experimental variable. The control group is not. Both the groups are measured after the experimental variable is introduced in the experimental group. The variation found/not found on the experimental group will determine the relationship between the two variables. (either as positive or negative).

## b) Before-After design

In this design, both the control and the experimental groups are measured before the exposure of the experimental group to the experimental variable. Then, the variable is introduced. The measurement is done for both the groups after the introduction of the variable. This ensures the reliability of the experiment.

Experimental design covers a few more types of experiments that are more elaborate and aim to determine the cause-effect relationship more specifically and precisely.

## Basic outline of experiment

In the experimental design, two groups are selected that do not differ from each other significantly. The 'experimental group' is exposed to the assumed causal variable. The 'control group' is not exposed to it. After experiment is over, these two groups are compared in terms of the assumed effect. Three types of evidences are required for testing causal hypothesis.

a) The first type is known as concomitant variation. The investigator could determine whether the assumed effect occurs more frequently among subjects who have been exposed to the assumed cause than among those who have not been exposed to it.

b) The second one is that 'y' did not occur before 'x' occurs since we are choosing two groups and apply a variable to ascertain its effect on one group. This being the case, the difference in these two groups after exposure of one of them to 'x' can be said to have decisively followed 'x'.

c) Evidence of third type is concerned with ruling out other factors as the possible determining conditions of 'y', which may be secured in several ways in an experiment. Such possible conditions are : a) eliminating factors that have occurred in the past b) avoiding contemporaneous events other than the exposure to the experimental variable c) maturational or developmental changes in the subjects of the experiment and d) the influence of the measurement process itself. Different procedures have been evolved to eliminate each of the above types of factors as possible determining factors of 'y'. One such measure is to select the two groups significantly matching with each other.

For easier understanding, the experimental design is explained in the following manner.

**Experiments**

**After only**
(Measurement of dependent variable only after the exposure of groups to experimental variable)

**Before After**
(Measurement of dependent variable before and after exposure of group to experimental variable)

Before After Single group

Before After with one control group

Before After with two control groups

Before After with three control groups

## Merits

In **After-only experiment,** the measurement is not done before the procedure starts. It has been assumed that both the groups are the same in all qualities. Once, the experiment is over, we do not know whether the difference is either due to initial difference or due to the exposure. Hence, in the **Before-After experiment,** the groups are measured bcth before and after the experiment. This design has certain advantages. They are :

a) They increase the sensitivity of the experiment.

b) They make it possible to determine the incidence of change in the dependent variable and to consider it in evaluating the result.

c) In case, the issue under study has a hypotheses specifying initially on the dependent variable, then, measurement is essential to test the hypotheses.

d) It increases the confidence that a difference found on after measurement is definitely due to the effect of experimental variables.

There is another design added to the above mentioned three designs which is referred to as **diagnostic design**. Diagnostic design adopts techniques to solve a problem. It goes a step ahead of other designs and finds the possible solution to a problem. The diagnosing starts with emergence of a crisis or a critical problem, finding out its causes, presenting of all ways and means of solving it and finally recommending of specific solution to the problem. Diagnostic design is used not only for clinical purposes but also to solve social problems.

Diagnosis corresponded to the fact finding aspect of clinical machine. Many of the problems which have been attached through field experimental have in fact been problems related to rather highly developed professional skills in such areas as human relations, training therapy, community organisation etc. For such problems, there are already available highly, skilled professional people who may be called upon to conduct the manipulation. Once, the basic role of relationships are worked out, there should be a collaborative diagnosis of the situation by the researcher and at least some part of the client organisation. The purpose of this diagnosis is to assess various factors that will be involved in executing the research design, the resistances that may by encountered, the dynamics of the situation, in regard to the problem of bringing about a change etc. Because, it is necessary to produce reactions and to bring about planned changes, the diagnosis must go beyond the mere statement of facts or the labelling of things as "good" or "bad"; it must move toward more causal thinking for it is only by manipulating its causes that we can vary dependent variable.

The next step should be joint planning based on an adequate diagnosis of the action that must be taken in order to manipulate the independent variable and to control other possibly confounding factors. To be most effective, this collaborative planning must go beyond the point of simply explaining to the client organization what is required in the research.

## Descriptive design

Many social researchers have been concerned with describing the characteristic features of communities or explaining an event or a phenomena. This sort of studies describe the factors relating to the issues like, social background, educational status and health status or enumerate different stages of an event. The researcher will collect and present the data with an elaborate description of the existing conditions, structure, customs and practices. There may be studies explaining the structure and function of organizations that can be grouped under descriptive design. Certain other social researches may have the objective of describing marriage or funeral practices of a community as a sociological concern. These types of research pre-supposes much prior knowledge of the problem to be investigated.

The methodology involved in this design is mostly qualitative in nature producing descriptive data on observed behaviour. The procedures to be used in a descriptive study must be carefully planned because the very aim of the research is to obtain accurate and complete information. Although, this design can make use of various techniques and procedures, they must be planned in advance even before starting the research work. The researcher should take every measure to avoid bias. He has to take into account the amount of work and the expenditure involved. The following are the steps to be followed in descriptive design.

## 1) Formulating the objectives of the study

The researcher has to define clearly the research issue that is taken for study. He must get a clear idea of what actually is his aim and why he is going to study and what aspects he has to explain. The objectives are to be specified with precision so that relevant data are collected to answer the questions raised.

When the objectives are clearly framed, then the stimuli for data collection could be precisely framed, so that there will not be any deviation or unnecessary data in the collection process.

## 2) Designing the methods of data collection

After formulating the objectives, the researcher must decide how he is going to collect the relevant data. The required data must be obtained with the help of a suitable technique. He must choose the one that is most appropriate for his own study from all the available methods of data collection like interview, observation or scaling techniques etc. In the case of descriptive design, more than one method of data collection could be adopted since the aim of this design is to ensure accuracy.

## 3) Selecting the sample

In descriptive studies, the researcher explains the behaviour pattern of mass of individuals in a group. It is not quite possible to collect information from all the people involved in that activity of the organization. Hence, the sample of the population should be studied. This sample should represent the total population in the study area. This sampling will help the researcher economise his study in terms of money, energy and time. If the population of the concerned area of study is less in size, the whole population may be included for the study. It is important to find out the total population and identify the unit of study.

## 4) Collecting and checking the data

In the process of data collection, the investigators should check for completeness, comprehensibility, consistency and reliability of the information collected. The validity of the tool (like the interview schedule) must be examined properly. The data must be checked for its completeness.

## 5) Analyzing the results

The investigator must decide at the beginning itself how the collected data has to be analysed. The investigator should avoid unnecessary labour like doing and redoing things pertaining to tables or addition and deletion of description. This will consume lot of valuable time. Such dispensable confusion can be avoided by planning these factors in advance. However, new ideas that occur

can be accommodated in the writing part. If any statistical computation is required, the data must be collected in such a way that it becomes amenable for qualification. While preparing the design itself, the researcher has to decide whether he is going to use computer or do the analysing manually. Perfect planning upto analysis should be made in designing the study.

## Exploratory or formulative design

In the case of exploratory studies, the researcher has minimum knowledge about the issue at hand and wants to explore the issue or situation in order to get more familiarity and information on it. Explanatory designs have the aim of formulating a problem for more precise investigation. Explaraotry studies help in developing hypothesis and also to clarify concepts. Sociological theories are often too specific or too general and for social scientists there is room for exploration of social events or issues with the background of these theories and their concepts. The investigator has to find out various causes responsible for it and also the most important variables. This type of exploratory research has to adopt certain procedures, so that the work could be carried out smoothly. They are discussed below.

### 1) Review of related literature

Having formulated the research topic, the investigator must review the work done on related topics by others. This will help him build upon his work more firmly and also help him to economize his efforts in the inquiry. He has to search for relevant materials in journals, books and abstracts. Some of the sources of information are Book Publishers, Bulletins and Journals, Documentation in Public Administration, Indian Agriculture Index, Indian Education Index, ICSSR Research Abstracts Quarterly, Indian Behavioural Science Abstract and Government Publication.

### 2) Experience Survey

Review of literature gives only the basic knowledge on the issue. It is the beginning of the formulative research. Written work is a small portion of the existing knowledge. There are people who have really experienced the situation or observed an event taking place. They are also rich sources of information, many people in the cause of their day to-day experience, by virtue of their particular placement as officials, social workers, professionals etc., are in a position to observe the effect of different policy actions and to relate these to problems of human welfare. These specialists acquire a fund of experience that can be of tremendous value in helping social scientists in their research analysis of the social problems in their study.

The respondents are carefully chosen on the premise that they will be able to contribute to the in-depth understanding of the study by providing valid ideas and rare insights. Efforts also should be made to select respondents to ensure a representation of different types of experience

in the study. After selecting the experienced respondents, it is necessary to conduct interviews with them. Unstructured interviews and structured interview schedules must be framed to conduct systematic interviews with the informants. It is essential to maintain a considerable degree of flexibility as the researcher is not fully familiar with the field.

### 3) Analysis of "Insight-Stimulating" cases

Other than the experience survey, the investigator should select a few important cases which would supply him with enormous amount of information, experience and stimulate insights. Sigmund Freud's insights to the psychological problems were stimulated by intensive study of his cases.

Stimulation of insights in the mind of the researcher depends on three other factors. Firstly, the attitude of the researcher must be highly favourable. He must be receptive and alert to acquire knowledge and find out facts. Secondly, the intensity of the study of the problem, the culture, community or the event may be examined thoroughly several times, taking note of all the concerned factors. Thirdly, getting insights rely on the integrative power of the researcher. The researcher should possess the capacity to draw together many diverse bits of information into an unified interpretation.

Exploratory design is to gain insight or to frame hypothesis. Only more carefully controlled studies are needed to test the general applicability of the hypothesis.

## Steps in Research Design

### 1. Selection of a problem and definition

The problem selected for the study should be clearly defined in operational terms so that the researcher knows positively what facts he is to look for, relevant to the study. Moreover, one's study should be delimited which will present accurate and precise results and convenient for data collection. The problem should be practicable in cost, of time and reliability of the results. Problems which are unwieldy and cannot be managed by the researcher should be avoided.

### 2. Data source

Three major sources of data., viz. 1) documentary sources both official and unofficial 2) personal sources and 3) library sources can be made use of according to the type of study. Data source in published and unpublished form can be used. Also primary and secondary source of data can be used. Data can be collected from primary source such as participant observation, personal interview, correspondence, questionnaire etc.

## 3. Nature of the study

The nature of the study, whether it is statistical, experimental or comparative type of study should be determined while planning for the design of the study.

## 4. Objectives of the study

Whether the design aims at a theoretical understanding of a problem or an exploratory effort to get clarity should be decided to have clarity in the study and to elicit good (response from the) respondents.

### Temporal context

The geographical limits of the design should also be referred to at this stage, so that the research related hypothesis may be applicable to particular social groups only.

### Dimension

The number of samples to be referred to is to be decided before hand to have a practical design. Also the limitations of the study on hand should be spelt out clearly in the research design.

### Socio-cultural context

A research design is always set to a context which has a social and cultural bearing on the individuals. The context and the conceptualization of the terms must be clearly spelt out in any research design. Each term used may have different connotations like economic, religious and political period of time. The period of time or the time dimensions of the study should be specified to have a clear exploration of the problem.

### Basis for selections

Time and cost factors form the basis for selecting data from the a sample form or sample enumeration. Studies drawn from the samples are more economical and efficient. However, the samples should be representative in character.

### Techniques of data collection

Techniques have to be selected suitably for collection of.required data. Techniques such as schedule, questionnaire, interview, observation or a combination of them may be chosen. The choice of the technique depends upon the accuracy of the data required are to be covered etc.

The researcher may select any one of the research designs for his study. The selection depends upon the purpose of the study.

The consideration which enters into the decisions regarding what, where, when, how much, by what means, constitute a plan of study or a research design. However, a research design has to be geared to the available time, energy and money.

It is stated that there is no such a thing as a single standardized research design. It is not a highly specific plan to be followed without deviations, but rather a series of guideposts to keep one going in the right direction. In other words, a study design is always tentative. As the study progresses, new conditions may come to light and it is necessary to change the plan as the situation warrants.

After the design is made, the research proposal is made and submitted. Initial draft proposal is subject to modification. Worthwhile research proposal will result only from a well designed proposal.

## Research proposal

A well designed proposal should have the following parts as components :

### 1) Title of the research

The title of research should be specific to the area of study. It should indicate the topic of study. The language used in the title should be professional, natural and simple. It should be as brief as possible.

### 2) Introduction

This refers to a brief explanation of the genesis of the problem. Here, the background material is to be given. The background or setting has to be given in a brief manner and this will be the starting point for the study.

### 3) Statement of the problem

Statement of the problem is an important aspect of the research proposal. If the problem is not carefully and scientifically stated, it will lead to confusion about the interest of the researcher. Use of clear, simple and concise statement is preferable. The major statement may be followed by minor statements. The problem can be stated either in question form or as a declarative statement.

### 4) Objectives

Objectives clarify the researcher's own mind and lead to more efficient enquiry. They give clear direction for achieving the ultimate goal. After stating the problem in a general form, the researcher may lay down the specific objectives.

## 5) Scope of the study

The scope of the study refers to the dissension of the study that is to be analysed. The scope depends upon four factors namely, (i) the availability of data. (ii) availability of funds. (iii) availability of time and (iv) the co-operation of the respondents.

## 6) Significance of the study

This refers to know, how the study will influence theory and practice. If the findings of the study helps to build scientific theories or help to practically solve problems, we can say that the study is significant. A study will have significance, if it can fill up gaps in research.

## 7) Review of related literature

It is customary to give in the research proposal a brief account of the important studies that are reviewed. The brief account of the review must deal with the following aspects of the studies reviewed :

(i) The problem that was investigated.

(ii) Design used (Experimental or non-experimental)

(iii) Research Techniques and tools used.

(iv) Population studied.

(v) Variables analysed

(vi) Pitfalls found in the studies

(vii) Chief findings that have emerged and the

(viii) Recommendations made.

## 8) Hypotehses

Hypothesis are formulated before the data are collected. The hypotheses formulated with limited evidence should be given in the proposal. It is these hypotheses which will be tested after the data are collected.

## 9) Definition, and operational definition

The concepts must be defined theoretically and operationally. Operational definition refers to stating the problem in observable and measurable terms.

## 10) Methodology

Here the methods used for data collection and data analysis, the tools and techniques used and their reliability and validity should be explained briefly.

## 11) Sampling

The type of sample that are to be selected and used and other details related to it must be given.

## 12) Research Report

Research report is an important aspect of research activity. It should be clearly stated as to how the report is to be presented. That is, the format of the research report should be given. (The number of chapters, the content etc.)

## 13) A Time schedule

A time schedule for the entire research activity to be undertaken should be given. It should have reference to the following aspects :

(i) Getting the theoretical background

(ii) Preparation of tools

(iii) Data collection

(iv) Processing and Analysis

(v) Writing the Report

## 14) Budget

A financial statement must be prepared and produced here. With regard to the expenditure, one may have to incur them on the following items.

(i) Stationery

(ii) Printing

(iii) Sample Selection

(iv) Field work

(v) Mailing and

(vi) Report preparation.

## Advantages of research design

The preparation of research design has the following advantages.

1. It saves researchers time, energy and money.
2. It helps to prepare and execute the various activities systematically.

3. It helps for the better documentation of the research activities.

4. It ensures a proper time schedule for the project.

5. It provides confidence and hope for the researcher to complete the work.

6. It provides satisfaction and sense of success at the completion of every state of the work.

## Social Surveys

The term social survey indicates the study of social phenomena through survey methods. Some of the definitions of 'social survey' are as follows.

1. The social survey is a comparative undertaking which applies scientific method to the study and treatment of current related social problems and conditions having definite geographic limit and bearing, plus such a spreading of facts, conclusions and recommendations as will make them as far as possible the common knowledge of the community and a force for intelligence co-ordinated action. - Shelly M. Harrison

2. The survey is briefly a method of analysis in scientific and orderly form for defined purpose of given situation or problem or population.

3. The sociologist should look upon surveys as a way and a supremely useful one of exploring the field, of collecting data around as well as directly on the subject of study so that the problem is brought into focus and points worth studying are suggested. - C.A. Moser

Following conclusions may be drawn from the above definitions regarding the characteristics of social survey.

1. Social survey is confined to the study of immediate problems of society eg. poverty, unemployment, delinquency, crime etc.

2. Its geographical scope is sufficiently limited and its field of study is geographically localised.

3. The purpose of survey is to prepare constructive programme of social research or removal of immediate evils.

4. The facts collected in a survey may form the basis of further social research on the matter. They may result in formulation of a new hypothesis requiring further elaboration, but it is not always the object of a survey.

5. Other necessary characteristics as mentioned by Harrison are 'co-operative effort' and use of 'scientific methods'. These however deal with the methodology and are equally applicable even to the social research.

## Difference between social survey and social research

Social survey and social research are identical in nature in many respects. Both of them deal with social phenomena. The methods and the techniques used in both of them are almost similar. Both are interdependent and are influenced by each other. Yet there are some fundamental differences between the two which may be classified as under :

| Social Survey | Social Research |
|---|---|
| 1. It is concerned with specific persons, specific problems and situations eg. prevalence of alcoholism among washermen in Madurai. | 1. It is concerned with general and abstract problems eg. effect of family environment upon delinquency. |
| 2. The object is to fulfil immediate needs and use knowledge available at a given time. It is thus practical in nature. | 2. The object is long time research of broad perspectives in order to develop more accurate procedures and theories. Thus its primary aim is theoretical in nature. |
| 3. The purpose is to improve the lot of men. It is thus utilitarian in nature. | 3. The purpose is to increase the general knowledge of men or devise improvement in the technique of study. It is thus purely scientific in nature. |
| 4. It results in a social reform, an administrative change or a remedial measure for removing immediate evils. | 4. It results in formulation of new theories or discovery of new techniques of study of modification of old concepts. |
| 5. It may form the basis of some hypothesis. | 5. It develops the hypothesis or thus evolves a theory. |
| 6. A hypothesis is not necessary for it. Generally a social survey is best undertaken without any hypothesis. | 6. A hypothesis is essential for proceeding on with social research. The social research is mainly concerned with the testing of hypothesis thus framed. |
| 7. It may be conducted on professional basis. Many surveys are conducted not for any interest in the topic but on payment from other interested parties. | 7. It is never conducted on a professional basis. No one stands to gain specially through such a type of study. Thirst for knowledge is the only incentive and its ultimate satisfaction is the only reward for a social scientist. |

## Object of a Social Survey

According to Moser, "A survey may be occasioned simply by a need for administration of facts on some aspect of public life, or be designed to investigate some cause and effect relationship or to throw fresh light on some sociological theory".

### 1. Supply of information on any problem

According to Moser, "the purpose of many surveys is simply to provide some one with information. Most of the surveys are, utilitarian in nature and are meant to provide information regarding practical problems. A survey may be carried on by the researcher himself, but there are also specialised bodies that carry on such surveys for others.

### 2. Description of a Phenomena

Surveys are also used for detailed description of phenomena. To a social scientist a survey may equally have a purely descriptive purpose as a way of studying social conditions, relationships and behaviour. No description of a phenomena would be accurate and complete unless we come face to face with it. Surveys help the researcher to come in direct contact with the phenomena under study and thus provide him with all the details that he needs. The purpose of these surveys is to collect general information and they are not meant to prove or disprove anything. These surveys may, serve as a basis for a hypothesis later on'.

### 3. Explanation of a Phenomena

Many enquiries, according to Moser are aimed to explain rather than to describe their function, may be theoretical to test some hypothesis suggested by sociological theory or severely practical to assess the influence of various factors which can be manipulated by public action, upon some phenomena but whichever be the case, the purpose is to explain the relationships between a number of variables. This is the theoretical importance of survey. Before we embark upon such surveys we must have a hypothesis to test, or a problem upon which we require some specific information to establish some causal relationship.

## Types of Surveys

Surveys may be classified into various types according to their subject matter, technique of data collection, regularity and so on. Following are the classifications.

### 1. General or Specific Surveys

When a survey is conducted for collecting general information about any population, institution or phenomena without any particular object or hypothesis, it is known as a general survey. Such surveys are mostly undertaken by the government for supplying regular data on many socio-economic problems. Census of population every tenth year is a typical example of such a survey.

Specific surveys are conducted for specific problems or for testing the validity of some theory or hypothesis. These surveys are naturally more pointed and only such information as is directly related to the specific purpose is collected. Information gathered through these surveys is generally of very little value outside the problem under study.

## 2. Regular and ad-hoc Surveys

Some surveys are regular in nature and must be repeated after regular intervals. For such surveys a permanent machinery for collecting information has to be set up. Most of such surveys are mainly economic surveys. Rural credit survey of family budgets can be cited as an example. Rural Credit Survey of the Reserve Bank of India which collects information regarding the condition of rural credit annually is also of the same type. Such surveys are undertaken where continuous data is required to study the trend or the effect of time upon the phenomena under study.

Adhoc surveys may also be conducted in phases. If the area of investigation is large, the whole survey being finished in two or more instalments. This would not entitle them to be termed as regular or repetitive type unless the same information is collected over and over again. Adhoc surveys are undertaken mostly for testing a hypothesis or supplementing some missing information regarding any research problem.

## 3. Preliminary and Final Surveys

A preliminary survey is generally known as 'pilot survey' and it is the forerunner of the final survey. The purpose of this survey is to get the first hand knowledge of the problem to be surveyed. It helps a person to get acquainted with the problem and the nature of respondents from whom the information is to be collected. It is therefore, very useful in preparing the schedule, questionnaire and or organizing the survey on proper lines. Final survey is made after the pilot study has been completed.

## 4. Census and Sample Surveys

In a census survey, survey of every single unit in the universe is to be conducted and information collected from it. In the case of a sample survey only a small part of it is taken as representative of the whole and data collected from it are made applicable to the whole universe. For example, in a study of smoking habit among the students at Madurai we may contact each one of the several thousand students to get the requisite information or a sample of few hundred may be taken and data so collected may be made applicable to the whole group. Sample surveys are becoming more popular these days because of their convenience, time saving and low cost. Besides this, techniques have been developed which permit the calculation of error with adequate precision.

## Planning and Execution of a social survey

Designing,organising and conducting a survey islike establishing a business enterprise. Both require technical knowledge and skill, administrative ability and specific experience or training in work similar to that being organised. Planning of social research programme is necessary for its successful execution. An unplanned research work may entail a lot of fruitless labour and wastage of time, money and effort. Most of the unplanned studies fizzle out before they have reached their final state because the researcher may find himself face to face with difficult problems which he had never thought of and for which he is never prepared. Patten has very aptly remarked - "only by carefully planning the survey from start to finish can reliance be placed upon results and in

many cases the findings ever reach the publication stage". A number of problems have to be considered in planning a social research project. Some of these are given below :

## 1. Selection of the Problem

The first and the foremost thing to be decided by the research worker is the selection of the research problem itself. Selection of an unsuitable problem would result in the inevitable failure. Following factors may be kept in mind while selecting the problem.

(a) The problem should be such, in which, the researcher may be deeply interested. His own interest will not only inspire him to work hard but also increase his power to grasp things quickly and keep his morale high in times of difficulties.

(b) The problem should be allied with the chain of thinking or research already in existence. Stray problems, if selected, become difficult to co-ordinate and do not add to the wholesome development of the theory.

(c) The problem selected need not necessarily be new. It may be an old problem or one on which work has already been done. There is a general temptation for the new research worker to hit upon some new problem, as it satisfies his sense of vanity, but from purely scientific point of view, verification of an old problem may be equally useful. Any hypothesis successfully proved in the repeat tests assumes the form of a perfect theory.

(d) The topic should be within manageable limits. It should not be too comprehensive. Here again a good deal of caution has to be taken. Lundberg has remarked "that social scientists tend to ignore painstaking study of limited problems and incline rather towards the grandoise method of philosophy of history".

(e) The topic should preferably possess some direct utility.

(f) The topic should be practically feasible. Practical feasibility conveys implied conditions. They relate to technique, time, money and other resources of the research worker. The topic selected should be such as it can be dealt with through existing techniques. Time, money and other resources at our disposal not only limit the dimension of our studies but also the nature of studies. All these factors must be thought out well in advance and a budget showing the proposed expenditure should be prepared before the research work is actually started.

## 2. Preliminary preparations

After the broad topic of study has been selected certain preliminary preparations are essential. The researcher must acquaint himself with all the material that is available on the matter. The research scholar should go through them. Allied matters should also be read. He should find out the problem, should contact the concerned people, discuss with them the problem, their findings, technique followed by them and the difficulties faced. This early preparation will make him more equipped and better suited to tackle the problem under study.

### 3. Selection of the Sample

Social phenomena is very vast. It is practically impossible to contact each and every body. Therefore, a sample is generally drawn from among the people about whom the study has to be made. There are different methods of drawing out a sample. The most essential feature of a good sample is its representativeness. The sample should include as fully as possible the qualities and attributes of the universe. It must be capable of yielding an accurate and reliable picture of the total. The sample must be simple and easy to be drawn. It should be mathematically possible to limit the errors. Above all, it must be practicable. An unpracticable sample, however excellent, is of no real use.

In order to draw a representative sample the 'universe' or the total population that has to be studied, must be known. A list of all cases, of universe, known as source list is thus prepared and the sample is drawn according to systematic sampling methods. Even after a sample has been drawn it should be tested by internal and external checks.

### 4. Deciding the method of study

Different kinds of methods are used in social research. Each method has its own advantage and limitations. The research worker should, therefore, decide in the very beginning as to what particular method he is going to use. In selecting the method he should take into consideration not only the suitability of the methods may be applied but a prior planning of the same is essential.

### 5. Nature and Information to be collected

Social phenomena is generally very complex and influenced by a number of variables. It is, therefore, essential to decide before hand what information is to be collected. Questionnaire and schedules are generally prepared for this purpose. They contain certain set of questions that are to be asked from the informants. Questionnaires and schedules bring uniformity and thus introduce objectivity in the study. A proper drafting of these documents is very essential. Any slip in the language may distort the meaning and may create in the replies. Information required should be complete. After the study has been completed, if some important information has been left out, it will require the same amount of labour and effort again to gather it. Therefore, enough care should be exercised in this respect.

Formal questionnaire may not be needed in some types of research work as case study, field observation, narrative, interview etc. But even in such cases the researcher has to make a plan, however short and sketchy to get an idea as to what information is to be collected. It may be in the form of interview guide, giving broad headings upon which questions to be put or only summary of the nature of information to be collected. Such a plan is sufficiently elastic and not fixed and unchangeable like schedule or questionnaire. The observer may make necessary changes in the matter to be observed as the research work is in progress and more experience is gained about the nature of the problem and the type of response available. In any case, some sort of previous planning is absolutely essential in every case.

## 6. Organisation of Survey

If it is a quantitative survey, a good deal of organisation work is required for successfully conducting it. A number of field workers may be required. If they are not trained they must be given necessary training. Proper arrangement should be made for checking and supervising their work so that they may not give false reports. After the report has been received, it has to be edited. A thorough checking of the forms has to be undertaken. After this, the data has to be classified, coded and presented in the form of tables. It is subjected to various kinds of analysis and ultimately conclusions are drawn proving or disproving the hypothesis. All this organisational work should be pre-planned. The researcher may put up his own organization or may take the help of some already existing professional body. If the project is quite big and likely to take enough time, it is generally more useful to set up one's own research organization. In short surveys, help may be taken from professional organizations. Because selection and training of the field workers is a time consuming process and may require a lot of preparatory expenditure. Even if a separate arrangement is made for this purpose, occasional guidance may be taken and consultations made whenever some special technique is involved.

## 7. Report

After the data have been analysed deductions are drawn. These generalisations are often put in the form of a report. If the survey has been undertaken at the instance of a third party care should be taken to see that all the information demanded by him has been supplied. At such times the surveyor may give his own opinion as well as the suggestions regarding the problem under study. The form of report to be submitted should be decided before hand as it is very helpful in deciding the nature of information to be collected and the form in which it is to be analysed.

## 8. Pretesting or pilot study

Even after a theoretical preparation has been made some sort of trial survey is necessary to gain specific knowledge of the subject and get an idea of the various problems likely to be faced in the course of the survey. These trial surveys are known as pilot study, pre-testing, test tube survey and so on.

In pre-testing, the same procedure is followed as is to be adopted in actual survey. Every detail is to be followed with the same precision. The number of cases to be included in the pilot study sample differs according to the nature of the survey.

In case any schedules or forms are to be used, they must be given a trial in the pilot study. The pilot study sample should be as far as possible representative of the main sample, in other words, it must include cases from all the different strata. Necessary amendments should be made in the questionnaire, method of study, the sample drawn and other plans of field organizations. If a substantial change has been introduced as a result of the pre-testing a sector or even a third pilot study may be essential.

## 9. Financial requirement

Every survey entails some financial expenditure which varies from survey to survey, according to the scope and size of the survey time required, nature of information to be collected etc. A prior

estimate of financial requirements is absolutely essential. Not only the expenditure required but also the resources through which the finances are to come should be pre-planned and definitely ascertained.

In preparing the cost budget all necessary items of expenditure may be taken into account. They may include such items as purchase of books, stationery, files, payment to be made to staff both clerical and field workers, conveyance and travelling expenses as well as other charges incidental to the survey work. A reserve of 10 percent for any contingencies must be made. Expenditure to be incurred on report must also be taken into account.

The general method is to divide the survey into following three parts, each part requiring almost equal expenditure.

1. Planning the survey and drafting schedule forms and questionnaire.
2. Investigation in the field, collecting data and editing schedules.
3. Coding, tabulating, analysing data and writing report.

## 10. Time Estimates

In any kind of planning, time factor plays a dominant role. The work has to be carried on according to time schedule. In fundamental research work time may not be an important factor and the researcher may take any period of time for completing his research, but in many types of short surveys specially those dealing with problems of immediate and transitory interest only, the time factor has a great importance and the survey must be completed within specified time if findings are to be of any interest to the people.

In preparing time schedule, pre-test surveys are very useful, but the pace of work in the beginning should be taken as standard. According to Parten, few people can continue to work at the pace established during the early days at a new job. Sickness, fatigue, holidays, extremely hot and cold weather, dismissals, errors, lack of supplies and many other factors tend to slow down the best of plans. A satisfactory time schedule takes into account as many of those contingencies as possible but it should also show time for the unforeseen.

No hard and fast rules can be made for calculating the time to be taken in a survey. This depends upon the nature of information to be collected, the number and efficiency of the field workers, co-operation of the respondents, length of questionnaire or schedule, the geographical distribution of the informants, arrangements for despatch, tabulation, analysis and printing of report etc. According to National Opinion Research Centre (N.O.R.C) of America, for an average size survey, six weeks to three months are required from the start of a regular survey until the data are ready for final form, two or three weeks are required before the last complete set is received and from one week to a month to tabulate the figures and compute percentages.

## Merits of Survey Method

Survey method has following merits in comparison to other methods.

1. A survey permits the researcher to come in direct contact with the people whom he wants to study. His inferences are, therefore, not based upon any theory or dogma but upon the facts of life.

2. This method permits greater objectivity. The use of large number of field workers removes the possibility of personal bias. In many cases actual survey has led to entirely different results than arrived at through purely theoretical wanderings.

3. General surveys bring to light a number of problems and propositions that would not have been possible by pure theoretical analysis. Analysis of survey results can point out to the hypothesis that may be completely outside the scope of existing theory. They may well said to be the introduction of new theory. For example, poverty was regarded as the cause of crime for a fairly long time, but increasing crime in the economically advanced countries have falsified this theory.

4. A survey allows full emotional impact of social situation specially when it is based upon participant observation. No amount of imagination can take the place of actual experience.

5. Survey method is one of the most widely used method. In case of social sciences where experimental method is not easily possible, survey method is the most scientific method for proving reliable data. Some of its techniques are equally used in Sociology, Anthropology, Psychology and natural sciences. The universality of application makes comparative study possible.

## Limitations of Survey Method

1. The first obvious drawback of survey method is the lack of money needed for conducting the surveys. In most cases the researcher has to recruit a large number of field workers, make arrangements for their training and supervision, arrange for the printing of schedules and the tabulation of collected data. All this require a considerable amount of money which very few people can afford. This is why surveys are now undertaken either by the government or through institutions aided by the government.

2. Survey is a prolonged and time consuming process. As a result, most of the surveys are conducted in a hurry and greater reliance is place upon whatever information is not always beyond question.

3. Survey techniques leave little scope for manoeuvering. The whole scheme is perfectly rigid and must be decided before hand. Any change in the major variables may mean conducting the survey afresh. A survey unless planned in advance may very well lead to complication later on, and howsoever perfect the pre-planning, it is impossible to visualise all the technical flaws that may be revealed only through experience.

4. The reliability and validity of the data collected through survey is not always without doubt. The reliability of the data to a great extent is dependent upon the honesty and efficiency of data to the field data to the field workers, the cooperation of the respondents, the adequacy and suitability of schedules etc. All these requirements are very seldom fully met and the collected data may contain much invalid and inaccurate information.

5. The study through survey method at a time becomes too much aggregative and therefore too much general. A survey must be fairly large to give valid result and the very extensiveness of the survey precludes the intensive study of individuals and even small group within the large population. To quote Camphell and Katona, "A sample survey designed to represent a population dispersed over a wide geographical area is likely not to give an adequate representation to any population characteristic which is highly localised. This means that the influence of specific local groups, for example, cannot be assessed through the usual national survey, since it is unlikely that more than a very few members of any such group would be caught in a cross section sample".

6. Survey method is not suitable for problems that require the study in the historical perspective. Ordinarily survey reporting refers to specific point in time or to a relatively short time period. Studies of origins and long term developments require research methods of more longitudinal character.

7. Most of the surveys are conducted on sample basis and the information collected is subject to sampling error. Besides this, the sample itself may not be representative, and then whatever information is collected may become unrepresentative of the group to which it is intended to be applied.

8. Survey method lays too much emphasis upon problems of immediate importance. It does not permit a more comprehensive study of the society. The study of the fundamental points of society in its wider perspective over time is not possible through this method. That is why most of the surveys deal with problems of immediate importance only.

## Model Questions

1. Discuss the role of research design in carrying out a research project.
2. What are the requirements of research design for descriptive and exploratory studies?
2. Elaborate and explain experimental design and its types.
4. What are the features of a good design?
5. What are the objectives of a social survey?
6. How do you plan and execute a social survey?
7. Discuss the merits and demerits of a social survey.

## References


1) Gosh B.N., Scientific Methods and Social Research, New Delhi : Sterling Publishers Private Limited., 1982

2) Goode J. Williams and Hatt K. Paul, Methods in Social Research, Tokyo : Mc Graw - Hill Kogakusha Ltd., 1952.

3) Sharma B.A., V. Ravindra Prasad D.R., and Satyanarayana P., Research Methods in Social Sciences, New Delhi, Sterling Publishers, 1985

4) R.N. Sharma and R.K. Sharma : Research Methods in Social Sciences, Media Promoters & Publishers Pvt. Ltd., Bombay, 1983, [Chapters 10, and 17].

5) Ram Nath Sharma : Techniques of Social Survey and Social Research, Rajhans Mandir, Meerut [Chapter 6]

6) B.A.V. Sharma, D. Ravindra Prasad and P. Satyanarayana : Research Methods in Social Sciences, Sterling Publishers Private Limited, New Delhi, 1984 [Chapter 19]


### LESSON - 3

# STAGES IN A RESEARCH PROJECT

Social research is the systematic method of discovering new facts, verifying old facts, their sequences, inter-relationships, causal explanations and the natural laws which govern them. There are three important stages in any research project. They are : 1) Collection of data, 2) Processing of data and 3) Report writing. These will now be considered in detail.

## Data Collection

Social research is based upon various types of information. These informations give knowledge concerning social phenomena. In order to carry on social research successfully, information should be gathered from proper sources. The more valid the source of information, the more reliable will be the information received which, in its turn, will lead to correct and reliable conclusions. Therefore, social research pre-supposes a knowledge of kinds and sources of information. Different types of social research require different types of information. This requires a wide knowledge of kinds and sources of information.

## Kinds of Information

Different kinds of information in social research can be classified into the following two types.

### 1. Primary Data

Primary data are the actual information which are received by the researcher from the actual field of research. These data are attained by means of questionnaires and schedules. In some

fields primary data are collected through interview and observation methods. The observation method, for collecting primary data, may be both participant and non-participant. Such data are known as primary because they are obtained by the researcher from the field of research directly and for the first time. Primary data are generally obtained through two sources. The primary source or primary data are the facts of living persons. The other source is the connected events known through observation such as the life of the group, traditions and customs and different aspects of daily life. Participant observation sometimes unravels such primary data which cannot be obtained.

## 2. Secondary Data

Secondary data are the information which are obtained indirectly. The researcher does not obtain them by himself . Such data are available generally from published and unpublished materials. Secondary data are gathered from information collected from the individuals and institutions through personal diaries, letters and survey documents. The secondary data again, are gathered through two types of sources. The first source is the personal documents such as diaries, letters, photographs etc. The other sources are the public documents such as books, manuscripts, records, census reports, reports of surveys by private institutions and various information published in newspapers and magazines. According to Lundberg, the information available from inscriptions on stones and objects of excavations may also be included in secondary data.

## Sources of Information

The sources of information are generally classified as primary and secondary. Different sociologists have classified the sources of informations differently. Of these, some more important views are as follows.

According to P.V. Young, the source of information can be classified into documentary source and field source. While the first includes books, manuscripts, diaries and letters, the second includes the information given by individuals.

According to G.A. Lundberg the sources of information may be classified as historical sources and field sources. The historical sources include sources representing the past incidents. For example, some inscriptions on stones of the ages of Ashoka the Great can be authentic sources of information about his time. Historical sources give information about the conditions and incidents which can not be known through observation or any other means of information. The historical sources include documents, papers and stone inscriptions concerning past, as well as the articles discovered through excavations. The field sources include informations received from concerned persons and also through the observation of their behaviour.

## Primary or Field Sources

On the basis of the above mentioned views the sources of information may be classified as primary or field sources and secondary or documentary sources. Primary sources are those from which information is gathered for the first time. This information is gathered by the researcher

himself. The sources of such information are the individuals and the incidents around them. Generally, primary source of information is gathered through direct observation, questionnaire, schedule and interview methods.

### 1. Direct observation

The chief primary source of information concerning social phenomenon is direct observation. This method requires that the social researcher should personally and directly absolve the conditions and incidents of his field of study. Direct observation is the most reliable method for gathering information concerning the life, status, conduct, behaviour, language, festivals, customs and traditions etc. Direct observation, again, has been classified into participant observation and non-participant observation. In participant observation, the observer lives in the groups or community as a member of it and participates in their life. On the contrary, in non-participant observation, the researcher does not participate in the group life but only observes as an external spectator.

### 2. Questionnaire

Questionnaire includes questions concerning different aspects of the subject of study. It is used in such cases where the subject of study is very wide and direct observation is not possible. It is also used about such things which cannot be known through direct observation such as the ideas, intentions and motives of the person concerned. Questionnaires may be sources of information only when the informers are well educated and prepared to co-operate with the research worker.

### 3. Schedule

Schedules are a particular type of questionnaire. The main distinction between questionnaires and schedules is that whereas in the former, the informers themselves reply to the questions. In schedules it is done by the researcher himself. He questions the informers and records the replies in the proper places specified in the schedules. The main advantage of a schedule is that it may be used even in the case of uneducated informer. However, it can be used only in a limited field, as compared to the questionnaire method.

### 4. Interview

In an interview, the researcher meets people and discusses his social problems with them. During the course of discussion, he gathers facts.

An interview is different from a schedule. A schedule includes some pre-determined questions asked by the researcher in a definite order without change. But the interview has no such definite form or order of questions. The researcher may ask any question on the basis of his insight into the problem.

## Secondary or Documentary Source

The sources of information through documents concerning individuals and institutions are known as secondary or documentary sources. The secondary sources represent social incidents, conditions and systems. Informations received through secondary sources are sufficient and useful.

They also provide information concerning the past which is not possible through any other sources. In order to gather complete information concerning some present social custom, it is generally necessary to know its historical background. The historical background can be known through secondary sources. Besides, secondary sources supply information concerning difficulties, precautions and methods of study. Documentary sources include books, survey scripts, memories, letters and diaries along with historical inscription and material published in news papers and magazines from time to time. To facilitate the study, secondary sources have been further classified into personal documents and public documents.

## a) Personal Documents

Personal documents include all the published and unpublished information documents by the individuals for different purposes. Personal documents are not written in scientific style not having an objective basis. They generally represent some ideas, values and feelings etc. In spite of being subjective and unscientific personal documents have been very much useful in social research. They provide important information about contemporary social circumstances, systems, customs, ways of life etc. The personal documents of notable persons are very useful sources of information concerning their social attitude. The causes of writing personal documents are generally as follows.

1) In order to prove the validity of some activity.
2) In order to accept one's mistake.
3) To achieve the satisfaction of ordered description.
4) To achieve the pleasure of literary presentation.
5) For research in personal documents.
6) To get relief from mental tension.
7) To attain riches, money and fame.
8) To fulfill some assigned job.
9) To provide clinical information and description.
10) To exhibit and express one's abilities and capacities.
11) To support some theory or ism.
12) Motivation for social service welfare.
13) To attain immortality.

Personal documents have been classified into four types - biographies, diaries, letters and memoirs.

**i) Biographies** - Biographies are generally autobiographies. Some great political leaders, social reformers and eminent persons write their autobiographies. These autobiographies are not mere collections of the incidents of their life. They provide self information concerning the social,

political, religious and cultural conditions and incidents of their time. Some autobiographies unravel information which could not be known from any other source. This is particularly true about the autobiographies of eminent political leaders. Few of such biographies are reliable, as most of them are written to achieve certain advantages.

Whatever may be the kind of biography, it is an important source of information about social conditions and incidents. Some biographers have elaborately described the evils of contemporary society and also suggested reforms. They also exhibit the social attitude of the subject of biography. An autobiography records the social philosophy of the author. The autobiographies are more simple than historical books.

**ii) Diaries** - Many educated persons have a habit of writing diaries concerning their social relationships and life incidents. Similarly, some persons daily write diaries to include programmes of the day. Then, some persons do not write diary daily but note only important incidents and deeper feelings and observations.

Dairies are written with different intentions. Some persons write diaries to remember important incidents of their life. Some persons write diaries to note things which they can not tell any one. This releases their mental tension. Some persons review their lives from time to time and note it in diaries. Diary writing is a confidential and private performance. Most of the diary writers have no intention to publish them. Therefore, generally it is expected that information recorded in diaries may be reliable. Hence, diary forms an important source of valid informations in social research as a personal document.

**iii) Letters** - Letters are often used by researchers as evidences. Letters are the medium of expression of the feelings, likes and dislikes, plans, attitudes, desires, emotions, ambitions and also important incidents of life. Letters are personal and very reliable. Therefore, they constitute one of the most reliable sources of information about interrelationships, such as friendship, love, marital affairs, family tensions and divorces etc.

**iv) Memoirs** - Some persons are interested in writing down memoirs of their travels, participation in bars, hunting expeditions, excavations, explorations and mountaineering assaults, etc. Such memoirs provide useful information in social research. Some historical memoirs have given very useful description of the contemporary time and circumstances. They have been acclaimed as very reliable historical documents concerning contemporary social, political and cultural conditions.

## Importance of Personal Documents

The above description concerning the importance of various types of personal documents, shows that they are important sources of information in social research. While providing information in objective conditions, they also throw important light upon the subjective conditions of the author. Therefore, apart from sociological data, they are very important sources of information for psychological analysis. Their information is relatively more reliable particularly when they have not been prepared with the intention of publication.

## b) Public or Official Documents

In the secondary source of information in social research, the public or official documents occupy an important place. A public document is the information gathered from government or non-government institutions. Public documents are of two types. Some documents generally remain unpublished such as of big companies, confidential records of government departments and non-governmental institutions. These are generally unpublished. Some other public documents are published in the general interest of the public. Such documents may be prepared by government or non-government departments. Some more important public documents are known as follows :

### 1. Records

Records occupy the most important place among public documents. Most of the governmental and non-governmental departments preserve so many types of records consisting of important information. For example, every police department keeps a record of its activity. Each police station keeps a record of criminals living within its area. Besides, private companies keep different kinds of social research. Information received through them have a high degree of reliability.

### 2. Published Data

Public documents include data published by government and non-government institutions from time to time. Such institutions conduct periodical survey concerning population, rate of mortality, birth rate, annual production, average annual income, foreign trade, marriages and divorces etc. A very important example of published data in India is - INDIA - A Reference Annual published by Government of India every year. It includes all types of data concerning almost all the departments of every part of Government of India and the States under it.

### 3. Journals and Magazines

Journals and magazines are important public documents including a wide variety of information which can be usefully utilised in social research. Most of these information are very much reliable. Letters to the editors published in various magazines and journals are an important source of information.

### 4. Newspapers

Newspapers publish news discussion on contemporary issues, reports of meetings and conferences, essays and articles, on living controversies and the letters of the readers to the editors. All these are important sources of information for different kinds of social research. The reliability of such information is very high, atleast in the case of renowned newspapers circulated among millions of persons. A comparative study of several newspapers, will give the investigator an insight into reliable and unreliable data. Newspapers in the countries under rigorous Government control and without much freedom of press, however, represent the views of party in power and therefore do not give much useful information. However, as the freedom of press in most of the democratic countries is respected, newspapers today supply very reliable information to social investigator.

## 5. Other Documents

Besides the above mentioned public documents, film, television, radio, and public speeches etc. are other important sources of information. They supply useful information about contemporary issues. The investigator, however, should be capable of sorting out the reliable material and distinguishing it from the unreliable material advanced by these sources.

## Utility of Secondary or Documentary Sources

The information received through the secondary sources are easy to obtain but difficult to use in social research. It requires a keen insight and a capacity to distinguish between reliable and unreliable data, to make use of documentary informations. Therefore, such information should be carefully judged before utilising it. This verification requires attention towards reliability and adequacy of facts.

### 1. Verification of reliability

Reliability of facts should be verified before using them in any social investigation. First of all, information should be obtained through the experience, honesty, ability, methods and tools of the investigator himself. Then the method of collection should be rigorously scrutinized. This method should be objective and scientific. It should be verified whether the sample made for social research is representative or not. Then, the organisation for collection of data should be scrutinized. The reliability of the methods and tools of study and the techniques of analysis should be verified. The sociometric scales utilised for measurement should be reliable and objective. After all this preliminary arrangement, the investigator should select some facts and verify them, so that he may have some idea about the reliability of total secondary data.

### 2. Verification of Adequacy

Besides reliability, the adequacy of facts should be judged before utilizing them in social research. In this connection, it should be remembered that whatever secondary data is collected should have an aim identical with the aim of social investigation. It is then alone that relevant data will be collected. Secondly, the units of available data should be verified so that adequacy of facts may be observed.

## Model Questions

1. Write a note on the primary data.
2. Discuss the importance of the personal documents.

## Books for further study


1. B.C. Tandon : Research Methodology in Social Sciences, Chaitanya Publishing House, Allahabad, 1979.
2. C.R. Kothari - Research Methodology, Methods and Application, Wiley Eastern Limited, New Delhi.
3. S.R. Bajpai - Methods of Social Survey and Research, Kitab Ghar, Kanpur.
4. R.N. Sharma and R.K. Sharma : Research Methods in Social Sciences, Media Promoters and Publishers Pvt. Ltd., Bombay, 1983.

LESSON - 4

# QUESTIONNAIRE AND SCHEDULE

## Questionnaire

### Definition

A questionnaire is a method in which a number of printed questions is used for collecting data. This list of questions is sent by mail to the respondents. After filling up the questionnaire they return it to the investigator. The questionnaire method has been defined by different sociologists in different ways. According to Bogardus, "a questionnaire is a list of questions sent to a number of persons to answer. It secures standardised results that can be tabulated and be treated statistically".

### Purpose of Questionnaire

i) To collect information from the respondents who are scattered in a vast area.

ii) To achieve success in collecting reliable and dependable data.

### Types of Questionnaire

P.V. Young has classified the questionnaires into two groups i.e.,

i) Structured questionnaire.

ii) Non-Structured questionnaire

A structured questionnaire contains definite, concrete and preordained questions. This type of questionnaire is prepared in advance and not on the spot during the questioning period. The structured questionnaires are used in a wide range of projects. This is used to initiate a formal enquiry and also to supplement and check data previously accumulated. These are mainly used in studies of economic and social problems, studies of administrative policies and changes, studies on the cost of living, consumer expenditures, public health and many other issues.

A non-structured questionnaire is used as a guide at the time of interview. In this method, the interviewer is free to arrange the form and timing of enquiry. Flexibility is the main advantage of this method. This method is applied to studies of family group cohesiveness, the studies of personal experiences, attitudes and the like.

The questionnaire is also divided into different types on the basis of nature of the questions. They are i) open ii) closed iii) mixed iv) pictorial. The open questionnaire is used in cases where new facts are to be searched out. This is mainly used for intensive studies of a limited number of cases. In this type, a respondent is free to express his views and ideas. The closed questionnaire is used when categorised data are required. Here, the informant chooses the answer from a set of provided responses. He has the liberty to express his own judgement.

The mixed questionnaire consists of both close and open type questionnaires. For social research, this method is very useful to initiate a formal enquiry and also to supplement and check data previously accumulated. These are mainly used in studies of economic and social problems, studies of administrative policies and changes, studies on the cost of living, consumer expenditures, public health and many other issues.

In the pictorial questionnaire, pictures are used to promote interest in answering questions. It is used extensively in studies of social attitudes and prejudice in children.

## Design of the questionnaire

The design of the questionnaire plays a vital role on obtaining the needed response. The following points may be borne in mind while designing the questionnaire.

1) The size of the questionnaire must be small and manageable. The respondent may lose interest in filling out a lengthy questionnaire.

2) It must have a neat get up. Special care should be taken with regard to margin, space between words and between sentences, printing and paper quality.

3) Questions should be clear. If there is no clarity in questions, the respondent may misinterpret them and supply wrong responses. This is likely to thwart the very purpose of research.

4) As far as possible, the questions must be interesting to the respondents. Lengthy questions may be avoided, for they are likely to create irritation and boredom on the part of the respondent. So questions should be brief and to the point.

5) Technical terms and abbreviations should not be used in the questionnaire. All respondents can not be expected to be familiar with them.. As a result such questions are likely to be left unanswered.

1) Know exactly what you want, so that you can ask the right question and phrase them properly.

2) Define terms that could be easily misunderstood.

3) Be careful in using descriptive adjectives that have not an agreed upon meaning.

## The mailed questionnaire

This type of questionnaire is a very useful technique in sociological research. When the researcher has to collect information from people who are scattered in different areas it is not possible for the researcher to go to all their areas to meet one or two of them. For instance, if a research aims to study the attitude of the IAS officers in Tamilnadu towards government policies, the researcher has to meet the IAS officers to collect data. But visiting all the places may not be feasible, specially when the time and funds are limited. He should alternatively mail the questionnaire he has prepared to economize his study. Its appropriateness will depend upon the requirements of the research problem with regard to 1) the type of information required 2) the type of respondents reached 3) the accessibility of the respondents 4) The precision of the hypothesis.

The mailed questionnaire should accompany an appeal to the respondent, explain the purposes of the research, who is sanctioning the study and a request to the respondent to respond and send it back. It should also contain every detail and directions of filling it up and guarantee anonymity of the respondent.

## Forms of Questionnaire

A numbers of facts are to be considered while designing a questionnaire.

They are :

1) The physical appearance of the questionnaire as it enhances the status of the study. The quality of paper and attractive printing creates confidence in the respondent about the research.

2) Clarity is important for the questionnaire because it is mailed. The choice of the words and spacing between questions should be borne in mind.

3) The size of the questionnaire must be small and manageable.

4) The questions should be arranged in proper sequence.

5) The use of technical terms and abbreviations must be avoided and as for as possible simple words must be used.

In sum, in a questionnaire, simple and familiar words must be used. Multiple meaning and ambiguous questions and words should be avoided. Leading questions which may create bias should be omitted. Also catch words, stereotypes and words with emotional bearing should not be used.

## Characteristics of a good questionnaire

The questionnaire is considered to be of high standard if it fulfills the following points.

1) Contain analytical questions

2) Indicate the topic at the top of the form

3) Clear and short

4) Limited number of questions

5) Necessary instructions to fill up the form

6) Clarity

7) Size of the questionnaire must be reasonable

8) Good presentation of the questionnaire

## Factors Affecting response

Different factors are responsible for varying degrees of response as stated below.

## Characteristics of the group

The response normally differs according to the characteristics of the respondent group. Generally educated people are more responsive as compared to illiterate people.

## Prestige of Sponsoring Body

If the institution has a good reputation, the percentage of response is very high.

## Importance of the problem

If the problem under study is important, a higher response is expected.

## Nature of Questionnaire

If the get up of the questionnaire is attractive, the respondents feel like responding quickly. Another important factor behind the response is that if the respondents are strongly in favour or against the concerned problem, a higher response is expected.

## Size of Questionnaire

If the questions are small in number, the percentage of response is high.

## Inducement for Response

Sometimes, the monetary inducement plays a vital role in getting a response. When the respondents get some inducement, he responds quickly. The inducement may be of many types, monetary or otherwise. In monetary inducement, some money is given to the people who are poor. Non-monetary inducement : a letter of appreciation, mentioning of the name of the respondent in the study and the like.

## Follow-up System

In this method, successive reminders are sent at regular intervals questioning the respondents to sent back the questionnaires as early as possible. Another method of ensuring replies is to sent the questionnaire through mediators.

## Reliability of Answers

The following factors affect the reliability of the response :

i) If the questions are ambiguous in nature, the respondents are not able to answer them correctly.

ii) The response through the questionnaire method may be unrepresentative in nature. The respondents who are not responding may belong to a particular class, resulting in the whole class going unrepresented.

iii) The questionnaire method is inapplicable to uneducated persons. Hence, social problems cannot be studied through this method.

## Test of Reliability

The reliability of the information obtained through questionnaire can be tested by the following ways :

i) The same questionnaire can be sent to the same person after a span of time. If the two replies are the same, the questionnaire can be said to be reliable.

ii) The same questionnaire can be tried on two similar samples. If the percentages of response are similar, the samples are to be regarded as reliable.

iii) If the majority of the questions are wrongly answered, then there must be something wrong either with the question or with the respondents, or with both.

iv) If the answers are mutually contradictory, the questionnaire is said to be unreliable.

## Advantages of the questionnaire technique

### i) Economical

The mailed questionnaire is popularly used when the geographical area covering the survey is very wide and the respondents are scattered over a vast area. So questionnaire method is more economical in terms of money, time and energy.

### ii) Anonymity

There are certain studies which require anonymity about the respondents. For example, the studies relating to marital relations and dreams etc. Moreover, it is expected that the respondents will feel free to give their opinions as they do not require to indicate their names on the questionnaires.

### iii) Uniformity

As the questionnaire is an impersonal technique, the uniformity can be provided by standardized wordings of questions. Moreover, this method places less pressure on the respondents for answering the questions.

## Disadvantages or limitations of the questionnaire

### 1) Limited in use

The questionnaire method of collection of data can be used only among the respondents who are well educated and ready to cooperate with the investigators. In other words, in the case of such studies relating to the common people, questionnaire method will be very limited in use and application.

### 2) Lower percentage of returns

In the case of mailed questionnaires, the proportion of returns of questionnaires with reply is very low.

### 3) Chances for misinterpretation

In case of mailed questionnaire, there are chances for the misinterpretation of questions and replies. So appropriate corrections cannot be done through repeating questions or by offering clarifications to the respondents.

## Pretesting and pilot study

Pre-testing is the process of an advance testing of the study design after the schedule or questionnaire has been prepared. It is very much useful and necessary to pre-test the schedule or questionnaire because it provides a test for the clarity of questions and correctness of the interpretations put upon them by the respondents. Moreover, pre-testing provides the possibility for the discovery of new aspects of the problem studied which are not anticipated in the planning stage. Thus, pre-testing is essentially a trial and error procedure wherein the successful trails are repeated and the errors avoided. Thus the object of pre-testing is to detect the discrepancies that have crept in and to remove them after necessary modifications in the schedule or questionnaire.

## Advantages of pretest

The advantages of pre-testing the schedule or questionnaire are as follows :

1) The pre-test reveals the strength and weakness of the questionnaire.

2) It helps to decide the type of questions to be adopted and best suited to the respondents.

3) The pre-test helps to resolve many problems of measurements. It will show how to develop proper line of enquiry in a logical manner in terms of respondents thinking. The sequential ordering or grouping of questions may be rearranged in order to seek their correct opinions or data or information for proper measurement and weighment.

4) The pre-testing also improves the design and effectiveness of the questionnaire in terms of format, quality, etc. The pre-test may also indicate the need for pictures, drawing, or physical objects to supplement the questionnaire.

5) Pre-testing provides the right estimate of resource required for completing the survey in terms of time, money, personnel and equipments to process the result successfully.

## Pilot study

A pilot study is a preliminary study conducted on a limited scale before the original studies are carried out to gain some primary information. Such preliminary survey of study of the universe enables to know the nature and various other aspects of the problem. The pilot study is undertaken on a small sample without formulating any hypothesis. However, the size and design of the pilot study is a matter of convenience time and money. After the pilot study is over, it is possible for the researcher to know whether the main study is still worth as planned.

## Advantages of pilot study

1) The pilot study helps in the selection of the respondents. It also exposes the inadequacies in the draft questionnaire. It also helps to know the non-response rate, reasons for it, desirable time for interviewing etc.

2) The pilot study provides the necessary training and motivations for the investigators to carry out their planned job.

3) The pilot study provides an opportunity for the investigator to learn about the local culture and builds up self-confidence for creating a report with the respondents. In other words, the pilot study provides a trial rehearsal to test the work of the interviewers.

4) The pilot study helps to identify the tools and equipments needs for the survey. The mode of transport and the tools and equipments needed for carrying the main survey have to be decided only on the basis of the experiences gained during the pilot study.

5) The pilot study helps to estimate the time and cost for completing the work successfully and also facilitates to effect economy wherever possible.

## Suggestions for its bet use

The questionnaire despite its limitations can be used as an appropriate data gathering device if the following points are borne in mind by the researcher.

1) The questionnaire must be properly framed. All the principles explained already must be adhered to while framing the questionnaire.

2) The size of the sample should be as large as possible.

3) While mailing the questionnaire to the respondents, postages for responses must be prepaid.

4) The questionnaire must be used only in cases where response is certain. This will prevent wastages.

## The Importance of Questionnaire Method

The questionnaire method occupies an important place in social research for various reasons.

1) The questions included in the list of questionnaire are standardised. The questions are real and create interest to the informants.

2) This is an indirect method of collection of data.

3) A respondent acquires some knowledge about many fields which were unknown to him.

4) The researcher or investigator does not have to make any administrative arrangements.

5) The method is economical.

6) A vast area and population can be studied easily.

## Schedule

## The Meaning of Schedule

A schedule is a list of questions which helps to collect data or requisite information. In this method, the investigator himself presents the questionnaire to the individuals whose responses

are needed. According to Goode and Hatt, "Schedule is the name usually applied to a set of questions which are asked and fill by an interviewer, in face-to-face situation with another".

Both questionnaires and schedules are very similar; but they also differ in some respects. A questionnaire is sent to the respondents by mail, whereas a schedule is used directly in interviews.

### Features of a Schedule

1) The list of questions is a mere document, so it need not be very attractive.
2) The schedule can be used in a limited area of research.
3) The schedule is put directly by the researcher and the answers are also noted down by him.

### Objectives of Schedule

P.V. Young has laid great importance on the following aims of the schedule.

### i) Delimitation of the topic

In the schedule method, the data should be collected in an objective manner.

### Types of Schedules

According to P.V. Young, the schedule can be divided into the following four parts :

### 1) Observation Schedules

These schedules contain some specific aspects on which the observer has to concentrate and collect information. With the help of these schedules, the observation becomes more accurate. This schedule serves many purposes ie. it is a standardising device; it is a specific memory tickler, and it makes observation more accurate.

### 2) Document Schedules

These schedules are used for recording data from case histories, documents, official records and so on. It is a very useful for collecting preliminary data. Tabulations are made from these schedules either manually or mechanically.

### 3) Evaluation Schedules

These schedules are used to gather information about some institutions or agencies. They help us to study their immediate problems.

### 4) Rating Schedules

These are mainly used in sociological or psychological research. They are especially applicable in cases where opinion, attitude and behaviour are to be measured. Through these schedules, different ranks or measures and rating is done.

## 5) . Interview Schedules

These schedules are used during interviews. Certain standard and specific questions are asked by the interviewer and he has to fill up all the information obtained in the table.

## Requisites of a good schedule

The following are the main characteristics of a good schedule, according to P.V. Young.

### 1) Accurate Communication

The basis of accurate communication is the proper wording of questions which are put to the respondents. The wording should be early understood by the respondent. Only then can accurate information be received. Technical terms and ambiguity must be avoided.

### 2) Accurate Response

To achieve this, the schedules must be kept short. It must have an attractive form and style. The questions should be simple and related to the concerned problem. Questions should be so constructed as to enable easy tabulation of the information gathered.

## Construction of a Schedule

For constructing a good schedule, the following steps should be considered :

1) The investigators should have proper knowledge about the problem and he should know what information is required for a valid and accurate generalisation on each problem.
2) For collecting exact information, the questions must be complete, lucid and precise. They should be so framed that the respondents can easily grasp their meaning.
3) The physical design of the schedule plays a vital role in getting the information quickly. The schedule should be well planned and good looking.
4) It should be short with proper margins.
5) The questions should be scientifically planned and cover all relevant aspects of the problem concerned.
6) In order to obtain valid information, the questions should be placed in a well-ordered serial.
7) After the schedule has been prepared, it should be tested on a sample population. If there is any shortcoming, it must be remedied and the schedule tested again.
8) Good quality paper should be used.
9) The print should be easy to read and well spaced.
10) If necessary, pictures may be used along with the questions to make the schedule attractive.

## Contents of Schedule

The schedule is divided into three parts according to the nature of the contents. These are :

## Introductory Part

The nature of the survey, the address of the surveyor, serial number of the case of interview, date and time of the interview and so on are mentioned clearly.

## Main Schedule

This consists of titles, columns and questions.

## Instructions

The reader or interviewer is given direction regarding the method of interview.

## Types of Questions

The questions of the schedule may be classified into following :

a) Open end question : The respondents are given freedom to express their views, as there is a wide range of choice.

b) Closed Questions : They do not allow the respondents to give answers freely.

c) Pictorial Questions : Pictures are drawn, and the respondent indicates the answer by selecting the picture he prefers.

d) Dichotomous Questions : Two alternatives are given : a positive one and a negative one.

e) Multiple choice questions : These questions consist of many answers. The respondent has to select any one of these.

f) Leading Questions : The reply is suggested in an indirect way. These types of questions create confusion. Thus as far as possible misleading questions should be avoided.

g) Ambiguous Questions : Questions which indicate alternate meanings and lack of clarity are called ambiguous questions.

h) Ranking Items of questions : Through these questions the preferences of the respondents are obtained.

## Defective Questions

A question is regarded defective if :

1) It is very lengthy.

2) If the nature of the question is indirect and ambiguous.

3) It is complex, and is regarding the personal life of a respondent.

4) It is suggestive.

5) A question that invites inaccuracy or unreliability.

6) It is superfluous to the problem at hand.

7) It is a question whose answers are universally accepted.

8) It contains abbreviations and technical terms.

9) It is addressed to the emotions of the respondent.

## Merits of schedule

1) In the schedule method, the answers are not biased as the field workers are personally present to remove any doubt.

2) The percentage of response is much higher.

3) A field worker is allowed to use, abbreviations of the answers. So it saves time.

4) The field worker is able to find out the sampling defect and can also remove it.

5) Since the field worker comes in personal contact with the respondent it enables him to probe more reply into the character, living conditions, general life style and other necessary details about the respondent.

6) The presence of the human element makes the situation attractive and interesting.

## Demerits of Schedule Method

1) It is an expensive affair and covers only small area.

2) It requires a large number of well-trained field workers, which in turn involves a great cost. Well-trained and experienced workers may not be easily available.

3) For conducting a survey, elaborate administrative and organizational arrangements are necessary.

4) In some cases, the presence of the field worker creates a source of bias in the interview, since the opinion of the respondent is able to be influenced by the field worker.

## Distinction between Schedule and Questionnaire

Apparently, the schedule and the questionnaire methods are similar. But there are also some differences between the two. They are as follows :

**Methodology :** The schedule is a direct method. Here, the researcher comes in direct contact with the respondent, whereas in the questionnaire method the data are collected indirectly through communications.

**Types of Question :** In the schedule method, the questions are very short. But the questionnaire method consists of rather included questions, to get detailed information.

**Reliability :** The information collected through the schedule method is more reliable. But in questionnaire method, reliability is some what doubtful.

**Area :** The schedule method covers only a limited area, whereas the questionnaire method covers a wide area.

**Clarification of Questions :** In schedule method, an investigator can collect the information from uneducated people by clarifying the meaning and purpose of the questions. But this is not possible in the questionnaire method.

**Collection of Information :** Through the schedule method, it is not possible to collect confidential information, whereas in questionnaire method, the respondent is free to express the actual fact.

**Use in Sampling Method :** Questionnaire method is not useful for sampling unlike the schedule, which is useful for sampling.

**Representativeness :** The data collected through the questionnaires are not completely representative, since the investigator is not present before the respondent who has to give the information. The respondent may be unable to understand the meaning of the questions, and hence may not be able to answer correctly, as required. This problem does not arise in the schedule method.

**Structure of Questions :** In the questionnaire method, the questions are made on the basis of the cultural and educational background of the respondents. In a word, it is respondent - oriented. But the schedule method is just opposite of this.

## Guidelines for Constructing Questionnaire / Schedule

The researcher must pay attention to the following points in constructing an appropriate and effective questionnaire or a schedule :

1) The researcher must keep in view the problem he has to study. It provides the starting point for developing the questionnaire / schedule. He must be clear about the various aspects of his research problem to be dealt with.

2) Appropriate form of questions depends on the nature of information sought, the sample respondents and the kind of analysis. The researcher must decide whether to use closed or open-ended question. Questions should be simple and must be constructed with a view to form a logical part of a well thought out tabulation plan. The units of enumeration should also be defined precisely to ensure accurate and full information.

3) Rough draft of the Questionnaire / Schedule be prepared giving due thought to the appropriate sequence of putting questions. Questionnaires or schedules previously drafted (if available) may be looked into at this stage.

4) Researcher must invariably re-examine and may revise the rough draft for a better one. Technical defects must be minutely scrutinised and removed.

5) Pilot study should be undertaken for pre-testing the questionnaire and it may be edited in the light of the results of the pilot study.

6) Questionnaire must contain simple but straight forward directions for the respondents so that they may not feel any difficulty in answering the questions.

## SPECIMEN OF A SCHEDULE OF AN ENQUIRY INTO EMPLOYMENT, INCOME AND SAVINGS USED BY THE TRAINEES OF THE FORD FOUNDATION (APPLIED SOCIAL SCIENCES RESEARCH UNIT, UNIVERSITY OF MADRAS)

**NAME OF INVESTIGATOR**

| Record of Calls | |
|---|---|
| Calls made<br>Date<br>Time and duration | |

### 1. Household

| | | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| 1.1 | Name | | | | | | |
| 1.2 | Age | | | | | | |
| 1.3 | Sex | | | | | | |
| 1.4 | Marital status | | | | | | |
| 1.5 | Address | | | | | | |
| 1.6 | Mother-tongue | | | | | | |
| 1.7 | Religion | | | | | | |
| 1.8 | Caste | | | | | | |

### 2. Education

| | | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| 2.1 | General | | | | | | |
| 2.2 | Technical | | | | | | |

## 3. Employment Status

| | | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| 3.1 | Employer | | | | | | |
| 3.2 | Employee | | | | | | |
| 3.3 | Own account worker | | | | | | |
| 3.4 | Family helper | | | | | | |
| 3.5 | No gainful employment | | | | | | |

## 4. Nature of Employment

| | | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| 4.1 | Industry | | | | | | |
| 4.2 | Occupation | | | | | | |
| 4.3 | Public or private | | | | | | |
| 4.4 | Secure or insecure | | | | | | |
| 4.5 | Regular (permanent) or casual | | | | | | |

## 5. Present Jobs*

| | | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| 5.1 | Working hours : normal for week days and Saturdays | | | | | | |
| 5.2 | Working hours : overtime | | | | | | |
| 5.3 | No. of days of work in the week | | | | | | |
| 5.4 | Duration of present employment | | | | | | |

*For rural areas see Appendix.

## 6. Previous Jobs

| | | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| 6.1 | No. held | | | | | | |
| 6.2 | Causes for leaving previous job | | | | | | |
| 6.3 | Salary in previous job | | | | | | |

## 7. Desire for Improvement in Conditions of Service

| | | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| In respect of | | | | | | | |
| 7.1 | Salary | | | | | | |
| 7.2 | Hours of work | | | | | | |
| 7.3 | Relationship with employer | | | | | | |
| 7.4 | Relationship with co-workers | | | | | | |
| 7.5 | Prospects | | | | | | |

## 8. Present Income

| | | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| 8.1 | Income from job (total) | | | | | | |
| 8.2 | Basic Salary | | | | | | |
| 8.3 | D.A. | | | | | | |
| 8.4 | Other allowances | | | | | | |
| 8.5 | Income from independent employment | | | | | | |
| 8.6 | Mode of payment : daily, weekly or monthly<br>Perquisites : kind or cash | | | | | | |
| 8.7 | Subsidiary employment*<br>Nature | | | | | | |

| | | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| 8.8 | Income | | | | | | |
| 8.9 | Days of work in the week | | | | | | |
| 8.10 | Hours of work in a day | | | | | | |
| 8.11 | Overtime work - nature | | | | | | |
| 8.12 | Overtime work - hours | | | | | | |
| 8.13 | Overtime work - earnings<br>Other incomes | | | | | | |
| 8.14 | Lands | | | | | | |
| 8.15 | House | | | | | | |
| 8.16 | Investments | | | | | | |
| 8.17 | Remittances | | | | | | |
| 8.18 | Others | | | | | | |

*Special attention of rural investigators is drawn to this detailed information is to be collected.

### 9. Present Unemployment

| | | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| 9.1 | Duration of unemployment<br>Since 1951 | | | | | | |
| 9.2 | Duration of present<br>unemployment | | | | | | |
| 9.3 | Sources of income, savings | | | | | | |
| 9.4 | Loans (Interest free) | | | | | | |
| 9.5 | Loans on interest | | | | | | |
| 9.6 | Interest on above | | | | | | |
| 9.7 | Duration of above | | | | | | |

| | | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| 9.8 | Pledging movables | | | | | | |
| 9.9 | Mortgage of immovable property | | | | | | |
| 9.10 | Sale of jewels | | | | | | |
| 9.11 | Sale of house | | | | | | |
| 9.12 | Sale of land | | | | | | |
| 9.13 | Sale of other property | | | | | | |
| 9.14 | Job sought - industry | | | | | | |
| 9.15 | Occupation | | | | | | |
| 9.16 | Salary expected | | | | | | |
| 9.17 | Willing to move to other parts in city | | | | | | |
| 9.18 | Willing to move to other districts | | | | | | |
| 9.19 | Willing to move to other states | | | | | | |
| 9.20 | Prepared or not prepared to train for a new job | | | | | | |

## 10. Casual Labour

| | | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| 10.1 | Days of employment per month | | | | | | |
| 10.2 | Wages per day | | | | | | |

## 11. Assets*

| | | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| 11.1 | House : Value | | | | | | |
| 11.2 | House : Income | | | | | | |
| 11.3 | Land : Value | | | | | | |

| | | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| 11.4 | Land : Income | | | | | | |
| 11.5 | Stocks & Shares : Value | | | | | | |
| 11.6 | Stocks & Shares : Income | | | | | | |
| 11.7 | Others : Value | | | | | | |
| 11.8 | Others : Income | | | | | | |
| 11.9 | Deposits : Value | | | | | | |
| 11.10 | Deposits : Income | | | | | | |
| 11.11. | Insurance : Value | | | | | | |
| 11.12 | Insurance : Premium | | | | | | |

*Mark H for hereditary and A for self acquired.

**12. General Reacgions to the Different Avenues of Investment**

| 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
| | | | | | |

**13. Savings**

| | | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| 13.1 | Total amount | | | | | | |
| 13.2 | Post office savings bank | | | | | | |
| 13.3 | Nidhis, etc. | | | | | | |
| 13.4 | Loans to others | | | | | | |
| 13.5 | Monthly rate of savings in 1994 | | | | | | |
| 13.6 | Monthly rate of savings in 1995 | | | | | | |
| 13.7 | Monthly rate of savings in 1996 | | | | | | |
| 13.8 | Monthly rate of savings in 1997 | | | | | | |

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 13.9 Are savings a regular and planned percentage of income? | | | | | | |
| 13.10 Why not larger savings : due to low margin? | | | | | | |
| 13.10 Due to security? | | | | | | |

## APPENDIX

(Additional Information to be Collected by Investigators in Rural Areas - Vide Block No.5 in Schedule)

| | Status of the Agriculturist 5.5 | Season when employed 5.6 | Employment for number of days in each season 5.7 |
|---|---|---|---|
| | 1 | 2 | 3 |
| 1 | | | |
| 2 | | | |
| 3 | | | |
| 4 | | | |
| 5 | | | |
| 6 | | | |

| | Employed for number of hours each day 5.8 | What kind of work is done in the off-season 5.9 | Number of days of employment in the off-season 5.10 |
|---|---|---|---|
| | 4 | 5 | 6 |
| 1 | | | |
| 2 | | | |
| 3 | | | |
| 4 | | | |
| 5 | | | |
| 6 | | | |

## Enquiry into employment, income and savings

## (May 1989) Guidance note

### Block 1 - household

Household means a group of persons, males and females, related or otherwise, having a common kitchen and pooling all or part of their earnings and incomes.

1.1 Begin with the head of the household. Head of the household is the person accepted as such by members to the household. (Give each member a Serial No. and follow this number throughout the Questionnaire).

1.2 Age means age at last birthday.

1.3 M - Male; F - Female

1.4 Marital Status :

M - Married

W - Widowed

D - Divorced

S - Separated

NM - Not married

(Indicate also the number of times by adding the relevant number).

### Block 2 - Education

2.1 General : Up to X standard : write the standard or form completed; beyond that indicate S.S.L.C., Inter, degree (specify).

2.2 Technical : Technical and professional degrees and diplomas. Also include technical training, recognized or otherwise, with years of experience after training.

### Block 3 - Employment status

3.1 Employer is one who is in business assisted by hired labour, part-time or full-time.

3.2 Employee is one who works for a wage or salary, in kind or cash.

3.3 Own account worker is an independent worker not in regular wage employment under any one and unassisted by paid labour.

3.4 Family helper is a person assisting the Head or any earner in the household in the performance of his craft, trade or professional work for which he is not paid.

3.1 to 3.4 - Tick off the relevant column.

3.5 No gainful employment : Indicate code number in classification below :

| | |
|---|---|
| Household worker | 1 |
| Children and students | 2 |
| Old age and sickness | 3 |
| Unearned income | 4 |
| Remittance holders | 5 |
| No desire to work | 6 |
| Unemployed for the first time | 7 |
| Unemployed not for the first time | 8 |

## Block 4 - Nature of employment

4.1 & 4.2 Write the actual occupation and place where employed (to be coded later).

4.3 Public means employment by the State or a Statutory Corporation or local board.

4.4 Security is with reference to tenure of employment.

4.5 Regular means of assured employment for the month. (Rural trainees may use the term permanent).

## Bock 5 - Appendix : Meant for Rural Trainees

Status of the Agriculturist : Indicate Code Number :

| | |
|---|---|
| Mainly owner - cultivator | 1 |
| Mainly tenant | 2 |
| Non-owing labourer | 3 |
| Non-cultivating owner | 4 |

Season :

| | |
|---|---|
| Ploughing | 1 |
| Sowing and transplanting | 2 |
| Weeding | 3 |
| Harvesting | 4 |

Off-season :

The interval between any two seasons.

## Block 6 - Previous Jobs

6.2 & 6.3 Refer to the job held immediately previous to the present one

## Block 7 - Desire for improvement in conditions of service

Tick off the relevant columns; if possible given details.

## Block 8 - Present Income

8.1 to 8.4 Apply to members figuring in 3.2

8.5 Applies to members figuring in 3.1 and 3.3

8.6 & below Apply to all persons having income.

## Block 9 - Present Unemployed

(Applies only to those currently unemployed)

9. Should indicate sources of maintenance during the present unemployed period. Answer giving amounts drawn from each source for the previous month.

## Block 10 - Casual Labour

10.1 Number of days employed during the last month.

10.2 Wage obtained on the last day of employment.

## Block 11 - Assets

All figures to be rounded off to the nearest hundred. Figures less than Rs. 50, record the exact amount.

11.7 Others : Items to be specified. Omit consumer goods, durable or otherwise.

## Block 12 - General Reactions to avenues of investment

Select from the following and arrange according to priority of preference :

| | |
|---|---|
| Building | Insurance, National Savings |
| Agricultural Land | Others (specify) |
| Cattle | Business (specify) |
| Bank Deposit | Money lending |
| Chit Fund | |

## Model Questions

1. Discuss the various steps in the preparation of a questionnaire.
2. Describe the advantages and disadvantages of a questionnaire.
3. Discuss about pre-testing and pilot study.
4. Discuss the stages involved in the formulation of a schedule and its advantages and disadvantages.

## Books for further study


1. B.N. Ghosh : Scientific Methods and Social Research, Sterling Publishers Private Ltd; New Delhi, 1982.
2. T.S. Wilkinson and P.C. Bhandarkar : Methodology and Techniques of Social Research, Himalaya Publishing House, Bombay.
3. Ram Nath Sharma : Techniques of Social Survey and Social Research, Rajhans Prakashan Mandir, Meerut.
4. S.R. Bajpai : Methods of Social Survey and Research, Kitab Ghar, Kanpur.
5. M.H. Gopal : An Introduction to Research Procedure in Social Sciences, Asia Publishing House, Bombay, 1970.


# LESSON - 5

# INTERVIEW METHOD

## Meaning

The interview method is a kind of verbal technique for obtaining data. It is the most commonly used method of data collection in the study of human behavior. It is a direct method. According to P.V.Young, "Interview may be regarded as a systematic method by which a person enters more or less imaginatively into the life of a comparative stranger. "

## Objectives of Interview

1) Laboratory study of verbal behavioral pattern under given circumstances.
2) Securing information from the person who alone knows the subject or the matter Under these two headings, we can classify the objectives in the following way.
    1) Formulation hypothesis,
    2) Collecting information about unknown facts through personal contact.
    3) Collecting information about qualitative facts.
    4) Improving the method of observation
    5) Collection information about various problems in different circumstances.

## Types of Interviews

The interview may be divided into three types ie. (i) Quantitative interview, (ii) Qualitative interview (iii) Mixed interview.

In the Questionnaire interview, certain set of facts are collected about a large number of persons, as in census.

**A Qualitative Interview** consists of certain complex, serious and non-quantifiable subject matter. This interview is confined to exploring the census of some events, sometimes it is called diagnostic interview.

**A Mixed Interview** is a method where both types of data are required. Some of the data may be quantifiable and some not.

## According to purpose

Every interview is done with certain aim or purpose. It may be classified into the following types :

i) **Clinical Interview :** The causes of certain abnormalities are ascertained.

ii) **Treatment Interview :** This is a signal to the clinical interview caused after the abnormalities are ascertained.

iii) **Selection Interview :** This interview is done to select cause of mental and visual with some particular qualities.

iv) **Interviews to fulfil Curiosity :** Through these interviews, the interviewer fulfils his own curiosity, which lurks in his mind.

v) **Research Interview :** This interview is done to collect information about certain problems to find out the truth.

## 3. According to Formalness

According to formalness, the interview can be divided into two main types : (1) Structured interview and (ii) Unstructured interview.

i) **Structured Interview :** A complete set of well-defined questions are used. In this method, highly standardized technique of recording are used. In structured interviews, generally even the alternative questions are fixed. Here the interviewer has to act according to the written instructions given in the schedule.

ii) **Unstructured Interview :** The interviewer does not follow a pre-planned list of questions. He enjoys full freedom to ask questions. He may revise the order of the questions to suit the needs of the respondents. In this type of interview a deep knowledge and skill on the part of the research are necessary. In the opinion of Johan Gautting, "the advantage of usntructured response is to be unprecise; that they permit the unexpected response."

## 4. According to Number

This category has been classified under two main types ie. (i) Group interview (ii) Individual interview.

i) **Group Interview :** A group of persons are interviewed for collecting information from them. This method will economise both time and money.

ii) **Individual Interview :** Here, a single individual is interviewed. Inter-personal contact between the interviewee can be established.

## 5. According to period of contact

There are two types of interviews on this basis ie., short contact interview and long contact interview. Sometimes, in social research, filling up of schedules or questionnaires, requires only a single sitting of small duration., For this type of job, a short contact interview is the most useful method, but where the case history of an individual in a group of persons is required, prolonged contact interview is necessary.

P.V. Young has laid emphasis on the classification of interviews according to the role assumed by the interviewer and respondent at the time of interview. These classifications which are mainly based on methodology, are given below :

i) Focused interview

ii) Repetitive interview

iii) Non-directive interview

## i) Focussed interview

This method has been applied by R.K.Merton for studying the psychological effect of Radio, Cinema, Television, etc. on the public. The main purpose of this type of interview is to examine a particular hypothesis. In such interview, the direction of the interview lies in the hands of the interviewer. Such interviews are based on pre-determined situations, gives importance to the emotional feelings or attitudes of the individuals in a particular situation.

## ii) Repetitive interview

The main objective is to study the dynamic functions, attitudes and behavior of certain individuals. For studying human behavior, this method is very useful. This type of interview requires that the respondents must be permanent residents of a particular locality so that they may be available for interview at any time. In a sense, it is a costly affair, since a permanent organisation has to be set up for this.

## iii) Non -Directed interview

There is no pre-planned set of questions, control or any direction in this method. An interviewer encourages the respondent to talk about the given topic without questioning him. For achieving

the proper goal, the interviewer should create a suitable atmosphere in which the individual is able to speak freely and fearlessly about himself. In problems like divorce and social disorganisation, this method is very useful. This is the method by which a list of information about underground activities is collected.

## Preparation for an interview

Before undertaking an interview, the interviewer must prepare the ground in a scientific manner. The stages of prior preparation for an interview are as follows:

1) The investigator must understand the problem properly.
2) Before taking an interview, interviewer must prepare the interview guide which gives an outline of the different aspects of the study. It may prove helpful in comparing data. It works as a "memory tickler".
3) An interviewer should approach an interviewee at a time when the person is free and in a relaxed mood.
4) The interviewer must seek a prior appointment with the respondents. If this is done, the interviewer may plan his own schedule accordingly.
5) Before undertaking an interview, an interviewer must gather the names, address and particular social habits of the respondents
6) When the general outline is prepared the cases which are to be interviewed have to be selected by various sampling methods.
7) Before starting the interview, the interviewer must give his full introduction to the respondent.
8) For getting accurate information, the head of the family or the leader of the group should interviewed.

## Technique (Process) of Interview

1) First of all, the interviewer must give the introduction about himself to the respondent.
2) After giving the introduction, the interviewer should explain the nature and purpose of the research project.
3) The interviewer should create an atmosphere in which the interviewee can express himself freely.
4) The interviewer must prompt the memory of the interviewee.
5) The questions should be put in a systematic way, in lucid language.
6) Emotional and leading questions should be avoided as far as possible.
7) At the time of interview, the interviewer must jot down certain important points between brief pauses, between two question or he can record them with a tape-recorder.

8) When the respondent loses interest in the interview, the interviewer should try to recapture his interest.

9) The interview should not be closed abruptly. The interviewer must observe the situation carefully and give a natural end to the discussion. The most important points should be summed up just before closing the interview.

10) After the interview is over the interviewer must write the report nicely and submit it as soon as possible.

## Validity of Interview

Some precaustions have to be taken to avoid the errors in interview. These precautions are :

1. The interviewer must examine the cause and the effect of the relevant aspects of the interview.
2. He should examine data from other sources also.
3. He must know about the relatives of the respondent and their attitudes, towards him (the respondent). This is important in the study of personal problems, attitudes and relations.
4. He must try to collect extensive facts about the important aspects.

If a researcher follows these precautions he may be able to draw an accurate conclusion.

## The qualities of a good Interview

A good interview technique depends upon the skill of the interviewer and proper framing of the questions. Interviewing is an art governed by certain scientific principles. Rapport should be established with the subjects. Their hesitation must be overcome and their confidence gained, so that they come out freely with their ideas and experiences. The aim of the interview should be explained clearly.

The interviewee should not feel any discomfort or tiredness while answering questions. The order of the questions may be changed to suit the emotive nature of the respondents. The interviewer should be free from bias and prejudices while interviewing a subject. While recording the responses two methods may be followed. Closed end questions can be answered by matching the appropriate alternative. When the questions are open ended the responses are to be recopied. So that at a later stage encoding and decoding becomes easier.

## Merits and Importance of the Interview

In social research, qualitative phenomena are not amenable to statistical analysis. In such cases, the interview is a very useful tool to gain insight. The following are the main advantages of interview method.

a) An interview is a means of getting direct knowledge personally and therefore, the information is reliable.

b) It is possible to study those phenomena which are not open to observation.

c) It is possible to study abstract factors like attitudes, feelings opinions reactions and so on.

d) It is possible to study the historical cases and past phenomena in an interview. The background of a case and the historical evidence can be best known eye-witnesses and the observers.

e) The background and the internal aspects of a phenomena can be classified in an interview.

f) One can learn the past, present and likely future behavior of human beings. The secret motivations, hidden desires and incentives working on human minds may be revealed by this method.

g) The inflow and outflow of information and knowledge may be mutually helpful to the interviewer as well as to the interviewee and barriers in the flow of information are easily eliminated.

h) The information given by the interviewee may be tested through cross-examination and emotional excesses and sentimental outbursts can be easily identified.

i) The method is highly flexible. New questions can be framed and cross checking can be done under this method.

j) Direct interview eliminates personal barriers, bring each other very close, making the study more fruitful, thus giving an opportunity to study the immediate reaction of the interviewee. The interviewer acts as a catalyst and he must learn to successfully handle delicate situations.

## Demerits and Limitations of interview

1) Interview yields subjective information which may or may not be true.

2) The method is not objective and scientific. It is difficult to verify the information.,

3) Personal factors emotions and sentiments may influence the interview and as such, the real issue cannot often be known. The data may be unreliable and invalid.

4) Even by the interview method, the correct information may not be elicited by the intervieweer. He may be biased, afraid, unwilling to reveal the truth or may be emotionally overridden.

5) The invertviewer may not be an expert either in the subject of study or in the art of interviewing or in both. In such a case an ideal result cannot be expected from the interview.

6) The method depends too much on memory and individual evaluation of the concerned problem and the pride and prejudice of the interviewee.

7) The method is time-consuming, costly and subjective.

8) The personal things and highly delicate matters may not be revealed by this method.

These limiations, however, can he greatly reduced by the careful, trained, educated and experienced interviewer. A successful interview requires a proper study design, a friendly atmosphere, a positive attitude and unbiased and objective minds. The success of the method greatly depends on the personal qualities of the interviewer. He must be intelligent, shrewd, tactful, charming, objective, trained, well-mannered, co-operative, witty and well-versed in human

psychology and intuition. According to P.V. Young, the limitations of the interview method can be overcome by :

a) Asking consistent and relevant questions.

b) Standardising the ways of asking questions.

c) Employing non-directive techniques.

d) Standardising ways of recording date.

e) Supplements interviews with projective techniques where a necessary (ie. the respondent is given an opportunity to press his attitude and mind on certain selected situations, materials or pictures.

The success of an interview depends to a great extent on successful communications between the interviewer and the interviewee. The establishment of a rapport between them is a must for a successful interview. It has to come in mind that an interview is an art.

## Guidelines for Successful Interviewing

Interviewing is an art. One learns it by experience. However, the following points may be kept in view by an interviewer of eliciting the desired information.

1) Interviewer must plan in advance and should fully know the problem under consideration. He must choose a suitable time and place so that the interviewee may be at least during the interview. Some knowledge of the daily routine of the interviewee is essential.

2) Interviewer's approach must be friendly and informal. Initially friendly greetings in accordance with the cultural pattern of the interviewee should be exchanged and then the purpose of the interview should be explained.

3) All possible efforts should be made to establish proper rapport with the interviewee; people are motivated to communicate when the atmosphere is favorable.

4) Interviewer must know that the ability to listen with understanding. Respect and curiosity is the gateway to communication and hence must act accordingly during the interview. For all this, the interviewer must be intelligent and must be a man with self-restraint and self-discipline.

5) To the extent possible there should be a free flowing interview and the questions must be well phrased in order to have full cooperation of the interviewee. But the interviewer must control the course of the interview in accordance with the objective of the study.

6) In case of big enquiries, where the task of collecting information is to be accomplished by several interviewers, there should be an interview guide to be observed by all so as to ensure reasonable uniformity in respect of all salient points in study.

## Causes of Errors in Interview

Having prepared the report, the interviewer has to recheck every thing to detect errors. It has to be tested for reliability so that one may go in drawing conclusions on its basis. following are the main causes which render an interview invalid.

1. Often interviewees, under emotional spells, embellish their accounts with high flown imagination and exaggerate their accounts in order to satisfy their varieties and create impression.

2. Sometimes there is communication gap between the interviewer and interviewee with the result that the interviewee says one thing and the interviewer understands something else.

3. Some interviewees deliberately try to mislead the interviewer. The interviewer must be mature and experienced enough to call off and rebuff such false interviewees.

4. Sometimes an inexperienced interviewer is offended by the behavior of interviewees and in a revengeful mood distorts the facts in his report.

## Means of avoiding errors in Interview

1. The interviewer tries to obtain extensive facts about important issues. This information works as test of reliability for the facts.

2. He examines data obtained through other sources.

3. Generally he tries to learn about the relations of the interviewee and his changing attitudes towards them.

4. He critically examines aspects like the relationship of cause and effect. This helps to determine whether the causes are always present or not when certain effects appear.

## Interview Guide

In Social Research, a printed or cyclostyled hand book of instructions for interview is made use of whenever the research is done through interview. According to P. V.Young the interview guide helps in the following four ways :

1. To focus attention on the main issues of the research project.

2. To maintain uniformity in interviews of different interviewers and thus help in the comparative study.

3. To collect same type of units for analysis and examination of hypothesis.

4. To accumulate specific concrete details as a basis for quantitative studies of life histories.

## The Place of Interviewer

The interviewer occupies a pivotal role. He gathers, tabulates and classifies the research material. The success of research depends largely on his proficiency and expertise. In the

interview, a personal relationship emerges between the interviewer and the interviewee, both affect each other. Therefore, the personality of the interviewer is of crucial significance. A man of charming manners and shrewd intelligence can easily create confidence in others.

It is the constant vigil and alertness of the interviewer which can ensure that facts about a case become known and that fake information is not given. It is again the skill of the interviewer which keeps the interviewee on track. As a rule, the interviewee tends to digress and worse still, indulges in irrelevant long talk just to satisfy his variety. Again, only a shrewd and sympathetic interviewer can study a problem. Thus, the interviewer is the centre from which all information disseminates. Therefore, he occupies a most vital role in the research by interview.

## Means of eliciting correct responses

According to Bogardus, in order to ensure correct responses the interviewee should be given maximum freedom of self expression and should be allowed to describe his reaction. There should be minimum interference while he is describing. His responses depend upon the skill and tactful approach of the interviewer and rules can be framed in this connection. Still the following points should be kept in mind.

### 1. Narrative Style

For allowing maximum opportunity of self expression to the interviewee, he should be allowed to tell his experience in a story form. That is, he may narrate his experiences in connected and sequential form.

### 2. Freedom of Description

The interviewee should be allowed to describe whatever he thinks worthwhile. Even if some irrelevant facts are being described, as long as he remains within reasonable limits of the main topic in question, he should not be pulled up.

### 3. Alert Direction

Though maximum freedom of self-expression is desirable, this can only be within the scope of the problem being discussed. If there is a total digression from the problem, the interviewee must be politely reminded to keep himself within bounds. This requires alterness and direction at the suitable occasions.

### 4. No harshness in direction

However, there should be no abruptness and harshness in providing the direction. Under no circumstances can an interviewer afford to offend the interviewee. Good humour is the essence of successful direction.

### 5 Interest hearing

The interveiwer must hear the interviewee with full interest. Nobody should be able to guess from his expression that he is bored or his mind is somewhere else.

## 6. Normal Emotional Attitude

Whatever may be the content of descriptions the interviewer must keep himself cool and remain self composed. a successful interviewer must possess a good deal more of self-restraint than others.

## 7. Winning the confidence of interviewees

For the narration of facts fully and frankly, the interviewee must have confidence in the interviewer. Therefore, the interviewer must try his best to win the confidence of his interviewees.

## 8. Appreciation of interviewee

If an interviewer can convince the interviewee that he appreciates his cooperation and greatly values the information given by him, the word of encouragement has a salutory effect on the interviewee who then gives correct responses.

## 9. Care in dealing with emotional issues

If the interviewer has to ask questions about the personal habits, likes and dislikes of the interviewee, he must do so skillfully. He must not appear to be interested in probing the secrets of the interviewees.

## 10. Objective attitude

The comments, whenever necessary, must smack of personal views, likes and dislikes of the interviewer. He must remain objective and should not let his personal feelings interfere. He must at all cost avoid criticism of the interviewee's behaviour.

## 11. Care for convenience of interviewee

This is, as a matter of fact, sine quo non of the whole process of interview. No body can hold an interview by issuing threats to interviewees or making them uncomfortable. In Social Research, the interviewee is an honoured person without whom research would come to a halt. Therefore, full care to his convenience has to be taken.

## 12. Adverse effect of quick report

Sometimes the interviewees get into high mood and begin narrating the adventures ad infinitum and ad nauseam. There is strong temptation under these circumstances to rebuke them and tell them to shut up. We cannot alienate an interviewee permanently, because he has lost his balance temporarily.

## Model Question

Discuss the types, techniques, merits and importance of the interview method.

## BOOKS FOR FURTHER STUDY


1. Young P.V. Scientific Social Surveys and Research.
2. Goode W J.and P.K.Hatt: Methods in Social Research.
3. Gopal M.H. An Introduction to Research Procedure in Social Sciences, Asia Publishing Housing, Bombay, 1970.

LESSON - 6

# EDITING, CLASSIFICATION AND TABULATION OF DATA

## EDITING THE DATA

Once data have been obtained either from primary or secondary sources, the next step is to edit the data, i.e.to scrutinize the same. The chief object of editing is to detect possible errors and irregularities. The task of editing is a highly specialized one and requires great care and attention. Negligence in this respect may render useless the findings of an otherwise valuable study. However, it should be noted that the work of editing data collected from internal records and published sources is relatively simple when compared to the data collected from a survey that needs extensive editing.

### Editing Primary Data

While editing primary data the following considerations need attention.

The data should be complete, consistent, accurate and homogeneous.

### Editing for completeness

The editor should see that each schedule and questionnaire is complete in all respects, i.e., answer to each and every question has been furnished. If some questions have not been answered and those questions are of vital importance the imformant should be contacted again either personally or through correspondence.

### 2. Editing for consistency

The editor should see that the answers are not contradictory in nature. If there are mutually contradictory answers, he should try to obtain the correct answers either by referring back the questionnaire or by contacting wherever possible, the informant in person.

### 3. Editing for accuracy

The reliability of conclusions depends basically on the correctness of information. It is necessary for the editor to see that the information is accurate in all respects. However, this is one of the most difficult tasks of the editor. If the inaccuracy is due to arithmetical errors, it can be easily detected and corrected. But if the cause of inaccuracy is faulty information supplied, it may be difficult to verify it e.g., inforamtion relating to income, age etc.

### 4. Editing for homogeneity

By homogeneity we mean the condition in which all the questions have been understood in the same sense. The editor must check all the questions for uniform interpretation. For example.

as to the questions of income if some informants have given monthly income, others annual income and still others weekly income or even daily income, no comparison can be made. Similarly, if some persons have given the basic income whereas others the total income, no comparison is possible. The editor should check up that the information supplied by the various people is homogeneous and uniform.

## Editing Secondary Data

Since secondary data are not collected directly by the investigator himself, care must be taken to ensure that the data compiled from published or unpublished sources are suitable, adequate and reliable. Such a data should be subjected to the consistency and homogeneity tests as in the case of primary data. Common sense and experience are chief guides of the editor in detecting errors in the secondary data on account of bias of the original investigator, size and selection of the sample wrong definitions and mistakes in calculations.

## CODING

Coding refers to assigning of number digits or letters or both to various responses whether quantitative or qualitative so as to enable tabulation of information.

The responses will then be recorded in codes and tabulated through the help of these codes only. Coding may be done before the designing of the returns. In the absence of pre-coding, coding may be done after editing but before the tabulation work has started.

Coding is all the more necessary in open or free response questions. In this case, a number of like responses are grouped into some distinct class categories. The codes are assigned only at the time of tabulation when the nature of responses has become known. It is necessary to ensure that each response fits in one category only.

The categories chosen should also be pertinent to the purpose. For example, in an inquiry in the defects in an automobile, the defect of ignition, braking body, etc., should be kept in separate categories irrespective of the way the responses are worded. The number of categories should be optimum. It should be neither too large to make the tabulation unwieldy nor too small to make the free response not so useful.

Coding in the modern age of computers is done after translating the information collected in the usual manner into the language of the machine. The most important thing in this is to transfer quantitative information into punch cards in decimal digits or the binary digits. The instructions for operations of tabulation, etc., have also to be prepared in the language of the machine in case of a computer.

A data processing language, known as COBOL (Common Business Oriented Language) was developed by the computer manufacturers in 1953. But now the most widely used language is the FORTRAN (Formula Translation) developed by the manufactures of the International Business Machines (IBM). However, there are machines to translate the information into the language of the machine and this work need no longer be done manually by the programmers. The only thing the programmers have to care is that the information collected is fit to be processed by the computers. After processing, the results can be decoded into the language it was collected. This also is done by the machine, designed for the purpose and therefore there is not much of a problem for the programmers.

## CLASSIFICATION OF DATA

### Definition of Classification

Once we observe the type or characteristic of the data, it is easy to rearrange the data into homogenous groups or classes having common characteristics. The process of the above type of arrangement is called 'Classification of Data'.

According to L.K.Connor, 'Classification is the process of arranging things (either actually or notionally) in groups or classes according to their resemblances and affinities and giving expression to the unity of attributes that may subsist amongst a diversity of individuals.

### Objects of Classification

Classification is the initial step in the presentation of any given data. The main objects of classification are the following :

a) to show points of likeness and unlikeness.

b) to bring into prominence the significant features of the data.

c) to save strain by discarding unnecessary details.

d) to obtain clarity and adaptability for manipulation.

e) to make comparisons easy.

f) to serve as a foundation for sound judgement and solid conclusions.

A probable question that arises at this stage is; how far should the classification be carried? No hard and fast rules can be laid down for this. It will depend upon the purpose in classifying the data. But in general, an ideal classification should possess the following three characteristics : ( i ) unambiguous ( ii ) stable ( iii ) flexible.

## Basis of Classification

Data can be classified according to the characteristic they possess. There are two types of classification : i) classification according to qualitative characteristic or attribute - this type of classification is called 'qualitative classification' or descriptive classification and ii) classification according to quantitative characteristic or variable. This type of classification is called 'quantitative classification' or 'numerical classification'.

## Qualitative Classification

In such a classification, the presence or absence of a certain attribute is studied. We form two groups for each attribute - one possessing the attribute and the other devoid of it. For example, a certain population can be divided into two groups as employed, unemployed, vaccinated, not vaccinated; literate, illiterate, male, female. These classes are exclusive of each other. Such a classification, where only two subdivisions are made in the study of one attribute is known as 'simple' 'two fold' or dichotomous' classification. Instead, the data can be classified into a number of classes depending on one attribute or two or more attributes. For instance, a certain population may be classified according to.

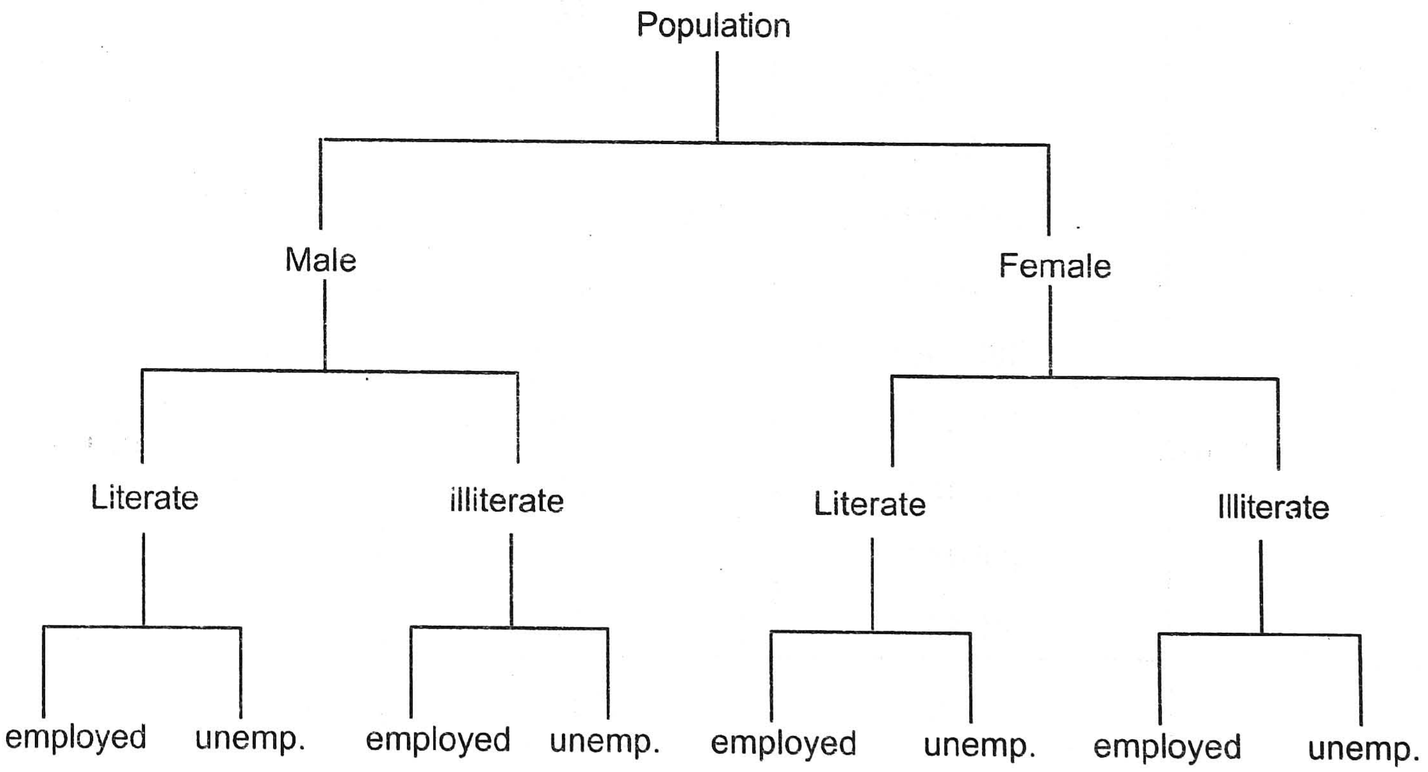
a) marital status as 'married, unmarried, widowed'.

b) religion as "Hindu, Christian, Muslim'.

c) employment, sex, literacy.

Such a classification is called 'Manifold Classification'.

The last classification c) is illustrated below.



**Quantitative Classification :** There are two methods of arrangement of variables in this type of classification. In the first method, we list all possible values of the variable, find the number of times each value occurs and give these two sets of umbers in a tabular form as shown in Example - 1 below. We call this Table 'Discrete Frequency Table' or 'Discrete Frequency Distribution'. In the second method, we group all possible values into a few classes, find the number of items in each class and put them in a tabular form. This table is called 'Continuous Frequency table' or 'Continuous Frequency Distribution'.

**Example - 1 : From a census report in a particular area.**

| Children in a family | No. of families |
|---|---|
| 0 | 10 |
| 1 | 40 |
| 2 | 42 |
| 3 | 30 |
| 4 | 10 |
| 5 | 6 |
| 6 | 2 |
| Total | 140 |

**Example - 2 : In a study of expenses of thousand students in a college**

| Expenses in Rs. | No. of Students |
|---|---|
| 50-60 | 80 |
| 60-70 | 90 |
| 70-80 | 100 |
| 80-90 | 200 |
| 90-100 | 300 |
| 100-110 | 65 |
| 110-120 | 55 |
| 120-130 | 50 |
| 130-140 | 40 |
| 140-150 | 20 |
| Total | 1000 |

## Statistical Series

Groups of items exhibiting some relation or comparison placed in systematic manner, constitute 'a statistical series'.

Classification of data can also be though of as different types of Statistical Series.

a) Time Series ( or Historical Series or Chronological Series)

b) Geographical Series (or Spatial Series)

c) Condition Series.

a) **Time Series :** Here, time is the predominant factor, according to which the data are arranged.

(eg.) If a company wants to know the trend of the sales over a number of years, the sales may be arranged in the following way.

| Years | Sales in lakhs of Rs. |
|---|---|
| 1970 | - |
| 1971 | - |
| 1972 | - |
| 1973 | - |
| 1974 | - |
| 1975 | - |
| 1976 | - |
| 1977 | - |

b) **Geographical Series :** Here, place or geographical location is the significant and deciding factor.

(eg.) Sales figures of a business concern for a particular year are arranged district wise in the following way:

| District | Sales in lakhs of Rs. |
|---|---|
| A | - |
| B | - |
| C | - |
| D | - |
| E | - |

If the concern is of national repute, the sales figures may be arranged state wise.

c) **Condition Series :** A series which is neither historical nor capital will be a condition series. In such a series, variation is size and amount is depicted. It is also known as 'frequency of occurrence at a given time and space.

(eg.) If the number of persons getting income within well- defined range are arranged, we call the series 'condition series or 'frequency series'.

| Incomes in Rs. | No. of Persons |
|---|---|
| 200-300 | - |
| 300-500 | - |
| 500-800 | - |
| 800-1000 | - |
| Total | - |

## TABULATION OF DATA

The collected material will be a confusing mass of data. In order to get a clear picture and to help further analysis these data have to be systematically arranged, summarized and presented in an intelligible form. So, first we note, whether the collected data related to 'Attributes' or 'Variables' and classify them. The classification of data is the preliminary step in the presentations of data.

Now, we see the different methods of presentation of data. They are the following :

a) Textual or descriptive, b) Semi-textual, c) Tabular, d) Diagrammatic and e) graphic. Last two forms of presentation-diagrammatic and graphic presentation of data are dealt in lessons 9 and 10.

a) **Textual form :** In this form, data may be given in a paragraph of text. Example : total rural population of India according to 1951 census is 2948 lakhs of which 2404 lakhs belong to agricultural classes. Of the total urban population of 618 lakhs, 531 lakhs belong to non-agricultural classes.

Of the rural agricultural classes, 687 lakhs are self supporting persons, 1404 lakhs are non-earning dependents and 303 lakhs are earning dependents. The rural non-agricultural population comprises 544 lakhs of which 1970 lakhs are self supporting persons. 326 lakhs are non-earning dependents and 48 lakhs are earning dependents.

This is not a very effective and impressive device for statistical presentation. In order to grasp the meaning of some of the points, one has to go through a lot of matter which takes up too much time. Here, data are generally unarranged and unsystematic. But here, emphasis may be placed on significant and important points.

b) **Semi-textual form :** it is a combination of the textual and tabular form. This method, though generally not used in practical work, may be used when a few figures are to be presented with some information.

An example of this form of presentation is given below :

Military training has become an important part of education in the various institutions of Uttar Pradesh. Government is also extending their fullest co-operation in organizing this. At present colleges in 19 cities are giving compulsory military training to their students. The number of colleges and cadets therein is given below:

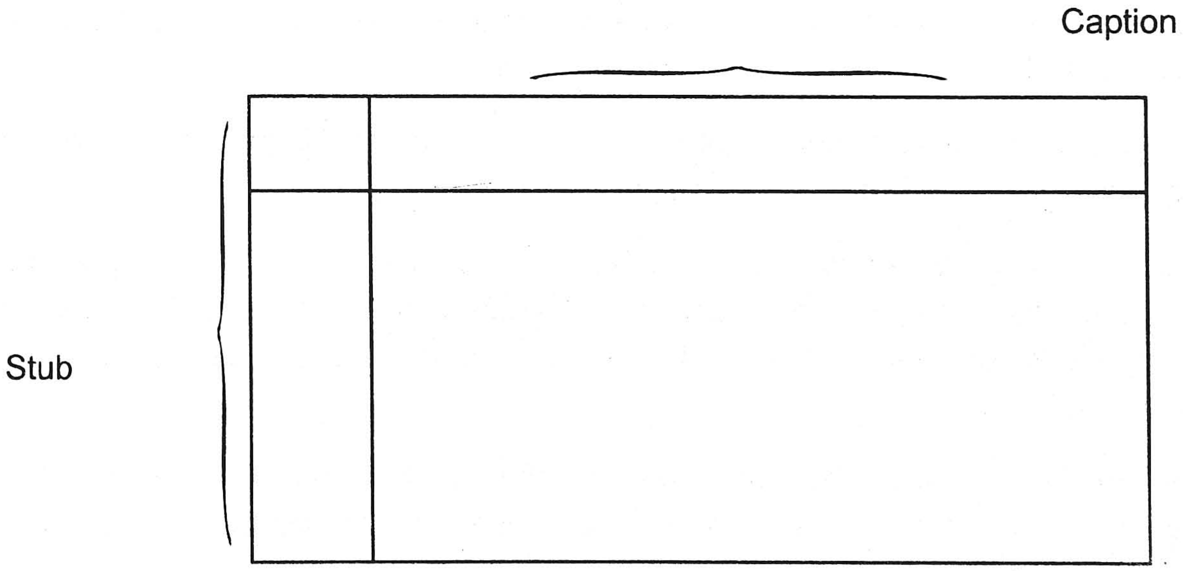
| | 1948-49 | 1952-53 |
|---|---|---|
| No.of Institutions | 65 | 140 |
| No.of cadets | 6,900 | 30,141 |

c) **Tabular form :** Tabulation is a process of orderly arrangement of classified data into series of boxes made up of rows and columns. The first column containing the headings for the rows is called the stub, and the headings of the other columns are called the captions, while the figures constitute the body of the table. Every table must be accompanied by a suitable title. Title, stub, captions and body are the four major parts of a table. Depending upon the specific case, there are other parts also. They are : (i) prefaratory note or explanatory note given just below the title (ii) footnotes given just below the body of the table and (iii) source note given in the end.

We have given below a schematic diagram showing the structure of a table :

**TITLE**

**Prefaratory Note**




**Foot Notes :**

**Source Note :**

## Construction of a Table

In order to construct an effective table the following points must be considered carefully.

1. **Title** : Each table should be given a clear and comprehensive title. The title should answer the questions :

(i) "What do the data represent?" (ii) "Where are the data form ?" (iii) "when" and (iv) "how classified?" A good title is always compact, yet complete.

Sometimes several lines may be required to write a title. These lines are usually written in the form of an 'inverted pyramid'. The following title illustrates this :

**EXPORTS AND IMPORTS THROUGH VARIOUS PORTS IN INDIA**

**1980-85**

2. **Stub and Caption :** Row headings and column headings must be brief and at the same time they must express exactly what is intended. Totals and sub-totals must be shown clearly, if necessary by using bold type for the relevant headings and figures or by light and heavy rulings. It is better to give the totals at the top or in the first column of the table, if only some of the components are shown and they do not add up to the total.

3. **Treatment of units :** Units of measurements of the figures given in the Table must be stated clearly at the head of each column or on the right hand side of the stub. If all the figures in the table are in the same unit, it is stated above the whole table.

4. **The body of the table :** The body contains the numerical data collected. It should be noted that sub-totals are given for each separate class of data and a grand total is given for all combined classes of numerical facts given in the body of the table.

5. **Prefaratory note :** A prefaratory note or head note is given just below the title. It is written in smaller or less prominent type of letters. It provides an explanation concerning the entire table or main parts of it; for instance, All data in million tons.

6. **Foot note :** A foot note clarifies some specific items or some specific part of the table or explains the commission thereof. It is placed just below the body of the table.

7. **Source note :** A source note should be complete and unambiguous. It should give the name of the author, title, time, volume, page, etc. The source note facilitates the reader in making reference to the original source to verify the quoted figures or to accumulate additional or detailed information.

8. **Arrangement of data :** Items should be arranged either in alphabetical, chronological or geographical order or according to size, importance or emphasis to facilitate comparison.

9. As far as possible figures should be arranged before tabulation. This would reduce unnecessary details. Large digits may be approximated and mentioned in thousands lakhs or millions.

10. Long unbroken columns of figures will be confusing. Hence, space should be left after every 5 or 10 rows.

11. If there are a number of columns in a table, they must be numbered or lettered for easy reference.

12. Figures to be compared should be given in columns which should be as near to each other as possible.

13. Use of ditto marks must be avoided. If a figure is repeated it must be shown each time. A ditto mark may be mistaken for the figure 11.

14. Abbreviations should be avoided, particularly in titles and headings. For instance, "Yr" should not be used for "Year".

15. If the data are very large, a single table drawn to represent these data would confuse the eye and lead to great difficulty in following the columns and rows at a glance. In such cases, the data can be represented in a number of tables. Each table should be complete in itself and should serve a particular purpose.

Some of the points given above are not universally applicable or universally accepted. In some cases, it will be impossible to follow them. Anyhow we must keep their purpose in mind. The above discussion may be aptly summarised in the words of J.C.Capt : "In the final analysis there are only two rules in tabular presentation that should be applied rigidly; First, the use of common sense when planning a table and second, the viewing of the proposed table from the standpoint of the user".

## Types of Tables

On the basis of the purpose for which they are constructed, tables may be classified into two types. They are

i) Reference or general purpose or primary tables.

ii) Text or special purpose or derivative tables.

**i) Reference table :** These tables present the original data for reference purpose. Hence they contain only absolute and actual figures and not round number of percentages. Tables given in census records and appendices of publications are of this types only.

**ii) Text tables :** These are usually constructed to present selected data from one or more general purpose tables. They bring out a specific point or answer a specific question. Texts are far smaller than general purpose tables. Ratios, percentages, averages and other computed measures are often included in text tables. Text tables are found in the body of the text.

On the basis of their content also we can classify the tables into two broad categories as follows :

i) Simple or one way tables

ii) Complex tables

**Simple tables :** A simple table contains data relating to only one characteristic. It gives information about one or more independent sets of questions. The following table is an example of a simple table;

**Example :**

**TABLE - I**

**Number of workers in a factory to age groups**

| Age group | No.of workers |
|---|---|
| below 10 | - |
| 10-20 | - |
| 20-30 | - |
| 30-40 | - |
| 40-50 | - |
| 50 & above | - |
| Total | - |

**Complex table :** A complex table presents data relating to two or more characteristics. It gives information about one or more sets of inter-related questions. It generally shows the relationship of one set of data to another data and are so arranged that comparison may be made between related facts. Depending upon the number of characteristics shown, complex tables may be classified as follows :

(i) Double or two -way tables or second order tables.

(ii) Treble or three-way tables or third order tables.

(iii) Manifold or higher order tables.

**(i) Double Tables :** Double tables are formed when either the stub or the caption is divided into two-co-ordinate parts.

**Example :**

**TABLE - II**

**Number of workers in a factory in different age-groups classified - sex-wise**

| Age-groups | No.of workers | | |
|---|---|---|---|
| | Males | Females | Total |
| below 10 | - | - | - |
| 10-20 | - | - | - |
| 20-30 | - | - | - |
| 30-40 | - | - | - |
| 40-50 | - | - | - |
| 50 & above | - | - | - |
| Total | - | - | - |

(iii) **Treble tables** : Treble tables are formed by making provision for three categories. The following example shows three characteristics (a) age -group, (b) sex and (c) skills.

**Example :**

**TABLE - III**

**Skilled and unskilled workers of a factory in different age-groups - sex-wise**

| Age-groups | No.of workers | | | | | | Total | | |
|---|---|---|---|---|---|---|---|---|---|
| | Males | | | Females | | | | | |
| | Skilled | Unskilled | Total | Skilled | Unskilled | Total | Skilled | Unskilled | Total |
| below 10<br>10-20<br>20-30<br>30-40<br>40-50<br>50 & above | | | | | | | | | |
| Total | | | | | | | | | |

(iii) **Manifold tables :** Where a number of attributes or variables are to be shown, the table becomes manifold. The following table gives classification of workers according to religion, age groups, sex and skill.

**TABLE - IV**

**Example :**

**Number of workers in factory according to Religion, Age group, Sex and Skills**

<table>
<tr><th rowspan="3">Religion</th><th rowspan="3">Age-groups</th><th colspan="6">No.of workers</th><th colspan="3" rowspan="2">Total</th></tr>
<tr><th colspan="3">Males</th><th colspan="3">Females</th></tr>
<tr><th>Skilled</th><th>Unskilled</th><th>Total</th><th>Skilled</th><th>Unskilled</th><th>Total</th><th>Skilled</th><th>Unskilled</th><th>Total</th></tr>
<tr><td rowspan="2">Hindus</td><td>below10<br>10-20<br>20-30<br>30-40<br>40-50<br>50 and above</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Total</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td rowspan="2">Muslims</td><td>below10<br>10-20<br>20-30<br>30-40<br>40-50<br>50 and above</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Total</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td rowspan="2">Others</td><td>below10<br>10-20<br>20-30<br>30-40<br>40-50<br>50 and above</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Total</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td></td><td>Total</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
</table>

## General rules for tabulation

The general rules for preparing the tables are

1) tables should suit the size of the paper

2) It should be systematically arranged.

3) Figures should be rounded up with proper foot notes explaining it.

4) Abbreviations or ditto marks should not be used.

## Use of Tables

1. The main use of tabulation is to enhance clarity and to give answers to various questions concerning the inquiry at the glance.

2. As the figures are presented in columns, it is easier to grasp the required information at one stretch.

3. Tabulation facilitates easy comparison by placing data side by side.

4. Tabulation eases addition, summarisation and detections of errors and omissions.

5. In a table, data are arranged in a logical order. Hence, the table helps the readers to have a logical sequence of idea of the data.

6. Tables avoid the necessity of repeating details and explanatory notes again and again.

7. Economy of space (while presenting) and economy of time (while reading) is effected by tables.

## Analysis of data

With regard to statistical application there are two types of analysis. One is descriptive and the other is inferential analysis.

Descriptive analysis limits generalization to the particular group of individuals observed. No conclusions are extended beyond this group and any similarity to those outside the group cannot be assumed. The data describes one group and that group only.

Inferential analyses always involves the process of sampling, and the selection of a small group drawn from a larger group. The small group is known as sample; the larger group, the population. In inferential analysis observations are made of the samples and based upon observations, conclusions are drawn about populations.

## Interpretation

The task of the researcher is not over just by presenting the data in tabular columns. He has to complete the analysis by presenting his findings in the form of empirical generalizations. He must do justice to his scientific pursuit by way of interpreting the data.

Interpretation is the process of stratifying a previous cross tabulation by another variable (called a test variable) under circumstances such that the has occurred (1) subsequent in time to the independent variable, but (2) Prior to the dependent variable. Such a procedure allows the further interpretation of an original relationship in terms of a third factor and Hatt.

Research interpretation is an effort to establish continuity in research through linking the results of one study with those of another. Secondly, interpretation leads to the establishment of explanatory concepts. Beneath what the researcher observes in the course of study, lies a set of factors and process which may explain whom he has observed in the empirical world.

Interpretation includes several other issues involving the relationship among the variables. They are (1) contingency and group differences, associative or functional (2) co-variational versus causal and (3) significant versus important. In theoretical inferences 1) the variables are empirically related. 2) the variables and concepts are related and 3) the concepts are related. In the case of policy inference, where it has policy implications, interpretations must be made with caution.

## Interpretation of results

Objective standards must be followed white interpreting research findings. The researchers must be well trained in the art of interpreting results. They must be trained to interpret results with utmost caution so as to enable an objective assessment. Objective interpretation of results gives understanding of the general factors which in turn can serve as a guide for further research studies.

## Pre -requisites for interpretation

The following are the pre-requisites which the researcher should bear in mind while interpreting the results.

### (i) Homogeneity of data

Data must be homogeneous. Homogeneity of data helps comparability of different situations. Suppose you want to collect data about poverty from different places and make a comparative analysis. If the data collected by you from different places are not homogeneous , your comparison will go wrong. To collect data that are homogeneous you must use the same definition of poverty in all the places from which you collect data. So homogeneity is to be ensured even at the time of data collection.

### (ii) Adequacy of data

Interpretation of data will be wrong and misleading if the data are inadequate. Suppose you examine the efficiency of two or three officers only in a company where there are forty officers and generalize their efficiency for the company officers as a whole, your generalization will be wrong or misleading. That is why it is said that for meaningful interpretation you must have adequate data.

### (iii) Appropriateness of data

The data that have been collected must be appropriate to the objectives the researcher wants to achieve. If the data collected are inappropriate, the results drawn will be wrong and

misleading. Even if the interpretation is objectively done, inappropriate data will result in biased or wrong interpretation.

## (iv) Scientific data analysis

It is not enough if the data collected are adequate and appropriate. Proper methods must be employed for analyzing the data. Wrong conclusions and incorrect interpretation will be made if the quality of the analysis is poor. Therefore, modern sophisticated, statistical techniques should be used for data analysis.

Analysis of data involves studying the tabulated material in order to determine the internet facts or meanings. It involves breaking down existing complex factors into simpler parts and putting the parts together in new arrangement for the purpose of interpretation. A preliminary analysis of the skeleton plan should be developed into a complete final analysis enlarged and re-worked as and when necessary. Larger divisions of material should be broken down into smaller units and re-arranged in new combination to discover new facts and relationships. Data should be studied from many angles as possible to find out new and newer facts.

The following are the four methods to start with analysis of data.

1. To think in terms of significant tables. That's possible from the data.
2. To examine carefully the statement of the problem and earlier analysis and to study the original records of the data.
3. To think about the problem in lay men's terms and to discuss the problems with others.
4. To analyse the data by using very simple statistical calculations.

## Interpretation

The statistical data and the information may be interpreted in various forms depending on the size and nature of data and the need of its interpretation.

Some of the common and important forms of interpretation may be described as,

### 1. Relationship

Relationship is one of the most fundamental basis to find out. Unless comparative analysis of study is made, a true form of relationship between subject and object can't be determined.

### 2. Proportion

Proportion is generally ascertained to determine the nature and form of changes in the form of study.

### 3. Percentage

If the object of enquiry is to determine, the nature and approximation only for the particular objective, then the method of percentage is should be chosen. Though the method of percentage is approximate and crude is often used in the sphere of calculating absolute figures.

## 4. Averages are other methods of comparison

These methods are used to interpret statistical data and information. If a long statistical table is to be analysed and interpreted, the help of various forms of measuring the central or other tendencies relating to them should be taken. The averages and other measures of comparisons are considered essential and form an integral part of the interpretation.

## Guidelines for interpretation

Interpretation of results must be done in a proper manner. The following guidelines will help researcher to interpret the results properly.

## 1. Make correct generalizations should be made

It is not always possible to conduct census. Researchers often study the characteristics of a sample. One will have to be very careful while making generalizations with the help of sample estimates. Suppose, you study the leadership behaviour of 100 women administrators in a particular organization. If you try to say that all women administrators in any organization would exhibit the same leadership; behavious then your generalization would be false. It would be false for two reasons.

(i) 100 women administrators are too small sample to make a generalization.

(ii) The sample you have studied constitutes the women administrators belonging to only one organization.

## 2. Sufficient care should be taken in drawing conclusions

Some researchers draw unwarranted conclusions. For example, in a study of administrative behaviour, it was found out that a manager suddenly stopped taking advice from his staff officer. The researcher interpreted the result by saying that the manager was not satisfied with the advice given by the staff officer. But the fact was that the manager had recently undergone a refresher course and hence he had developed confidence in himself. Thus the researcher made an unwarranted conclusion.

## 3. Fair and unbiased interpretation to be made

You may have bias, prejudices and inhibitions but do not allow them to get reflected in the interpretation of results. When two students have failed in the examination, one student thought that his life was doomed. He was a pessimist. Another student did not take his failure that way. He thought that it was a lesson for him to work hard thereafter. He took it as a challenge to active success in the next examination. He was an optimist. You may either be a pessimist or an optimist. But as a researcher you should never allow your pessimism or optimism to guide you to interpret the results. As an interpreter you must be unbiased, fair minded and interpret the results objectively.

## 4. Carefully chosen statistical concepts

One statistical concept which is often wrongly interpreted is the concept of coefficient correlation. In one study, it was found out that there was high degree of correlation between the scores of two subjects of a group of students. The researcher interpreted that the students who do well in one subject would do well in another subject it may be sometimes true of intelligent students. But with regard to average or below average students, such interpretation will not be good. So the researcher must use such statistical concepts with abundant caution.

The researcher has to go beyond the collection and analysis of data. He must also seek to show that his observations points to certain underlying relations and processes which are initially hidden. He has to show that his observation has a meaning much broader and deeper than the one that appears on the surface level. His task extends in interpreting the findings of his study to understand its real significance. Interpretation is the search for broader meaning of the research findings. The search involves viewing the research findings in the light of the other available knowledge, a theory or principle.

Interpretation has two aspects. Firstly, there is an effort to establish continuity in research through linking the result of a given study with those of another. It is through interpretation that the researcher can unravel or understand the abstract principle beneath the concrete observations. Then the researcher can easily proceed to link his findings with those of other studies conducted in diverse settings but reflecting the same abstract principle at the level of their findings. Based on the recognition of the abstract theoretical principles underlying his findings can make predictions about many events. Interpretation is conceived as a special aspect of analysis rather than a separate aspect.

## Model questions

1) Explain the process of classification, tabulation and analysis of data.

2) Discuss the process of analysis and interpretation of the data.

## References


1. Janathan Anderson, Berry H.Durston and Millicant Poole, Thesis and Assignment Writing, 1970.

3. Claire Sellitz, Lawrence Wrightsman and Stuart W Cook, Research Methods in Social Sciences, 1976.

3. Young P.V., Scientific Social Surveys and Research.

4. Goode W.J. and P.K. Hatt : Methods in Social Research.

5. Gopal M.H., An Introduction to Research Procedure in Social Sciences, Asis Publishing House, Bombay 1970.

## BOOKS FOR FURTHER STUDY

1. D.C.Sancheti and V.K.Kapoor : Statistics, Theory, Methods and Application, Sultan Chand and Sons, New Delhi, 1978 (Chapters II,III)

2. C.R.Kothari - Research Methodology, Methods and Techniques, Wiley Eastern Limited, New Delhi, (Chapter 6)

3. S.P.Gupta : Statistical Methods, Sultan Chand Sons, New Delhi, Chapter I.

4. Sadhu Singh : Introduction to Research Methodology in Social Sciences, Himalaya Publishing House, Bombay.

### LESSON - 7

# REPORT WRITING

Research reporting is the oral or written presentation of evidence and the findings in such detail and form as to be readily understood and assessed by the reader so as to enable him to verify the validity of the conclusions. Research report writing is the result of the research investigation. It is at the stage of reporting that the researcher assembles the findings of the study, draws conclusions and evaluates his own findings. Report writing is thus the end product of a research activity. It is a highly skilled work as interesting, fascinating and at the same time challenging, gruelling and sometimes even exasperating experience. Writing a research report is a technical activity which demands all the skills and patience from the researcher. It requires "considerable thought, effort, patience and penetration and an overall approach to the problem, data and analysis as well as firm control over language and greater objectivity"

To write the report, a vast amount of planning and presentation is necessary. Besides, perfection in the research report is achieved by continuous and persistent thought and creative and intelligent writing. Only hard and patient work on the fact, careful and critical assessment and intelligent planning of the organisation of the report can facilitate communication. There is no standard criterion for the organisation of a report popular or technical for, the form depends on each investigation, problem, the novelty or familiarity of the methods, nature and volume of facts, techniques of analysis and so on. But, there are a few guidelines. The report for the lay public should be simple, interest and lucid. Similarly the research report for the administrator should neither be too general nor too technical.

A research report is meant for several categories of audiences viz. scientists engaged in fundamental research, those engaged in applied research, practitioners, policy formulators, funding agents or sponsors and the general public. The report should focus on the target audience and include the material of interest to other audiences only if it does not interfere with the material required to the specific audience. The material that would not be appropriate for the target group can also be separated from the text and incorporated in the footnotes or appendix. Uncertainly about target group results in ineffective communication.

## Reflective thinking

Research is a thinking power. It is a mechanical process of gathering information and presenting them. Researching power of the researcher, is the efficient tool for research decision, interpretation and generalisation. According to Whitney There's a high positive correlation between good thinking and effective writing.

According to M.H.Gopal "Research is not merely the accumulation, evaluation and assimilation of facts. It's also and more so, re-building them into meaningful whole. This demands patient, deep and alert thinking". A researcher must develop his thinking process.

## Outlining

Outlining is the outcome of the thinking process, it's the next step in writing the research report.

Outlines are the framework upon which long written works are constructed. They're an aid to the logical organisation of the material and the reminder of the points to be emphasized in the report. Outlining of the research study is made up of two stages, namely beginning stage and the writing stage.

The outline prepared at the stage of writing should be elaborate and it should be prepared at three stages, namely

1. Topical
2. Paragraph
3. Sentence Outline.

Topical outline comprises chapters in broad aspects. Paragraph outline contains idea behind each paragraph. Sentence outline indicates the points to be included in sentences.

M.H.Gopal suggests the following points while planning an outline.

1) The outline should be as detailed as possible.
2) It should not be vague.
3) It should fulfill the consideration of chronology topical unity, coherence and transition.
4) Each paragraph must contain one idea.

# TYPES OF REPORTS

Before the preparation of the report, the researcher has to consider the potential guidance, the total research situation and their technical background and knowledge. The report may be short report or a long report.

At the beginning of the short report, there should be some brief statement concerning the authorisation of the study, the problem to be examined and its breadth and depth. Following this statement, a summary should come along with the conclusions and recommendations. Short reports are appropriate for studies in which the problem is well defined. Such studies of limited scope require only limited time personnel and money, and for it, methodologies are simple and straight forward.

The long report may be technical report or a popular report. The technical report, is prepared for a specialist who is interested in understanding the technical procedure and terminology used in the research project. The report will be in technical language.

The popular report is intended for persons who have limited interest in the technical aspects of the research methodology and research findings, the audience will include laymen and even top executives who want summery reports.

When the audience comprise of both technical and non-technical people it will be necessary to develop a composite report that serves both audiences. In such a situation appendices may be used to include technical data.

Research reports contain findings of the fact, analysis of the facts, interpretation, conclusions and recommendations. Based on the approach the reports may be divided under the following heads :

1) Business Report or Memorandum
2) Project Report
3) Dissertation
4) Commission's enquiry report
5) Thesis

A business report memorandum is simple business communication from top to bottom or from one functional area to another or any specific activity of day to day business activity.

A project report is a report of a project undertaken by a functional area or by any individual relating to business, industry or society.

A dissertation is a detailed report on the subject of the study undertaken. This is used as a document to be submitted for getting a higher degree from an university.

The two terms dissertation and thesis are used interchangable. A thesis is expected to make original contribution to knowledge. It's the culmination of substantial piece of original work, a period of atleast one year. A commission of enquiry report is a detailed report, prepared by an expert body or by a justice of enquiry. The report is the result of the commission's detailed and indepth enquiry relating to a specific matter or subject.

On the basis of presentation, reports my be classified under the four heads namely,

1. Inductive reports.
2. Deductive reports.
3. Step by step reports
4. Time sequence reports.

The inductive pattern or report writing, the oldest form of reporting goes from the specific facts, figures, materials of research to general conclusions, summary, recommendations and the like.

The deductive pattern of report is also known as executive, management or action getting report. Those reports go from the general to the specific, that is from more important matters to less important matters.

It's almost like a newspaper report. The executives want only the major findings and recommendations of the report. No voluminous details that are required for the technical reports are required here.

Step by step report is a report on the functions or operations, frequently following one another during a particular period of time and this kind of report is useful for the manufacturing process of the industrial sector.

Those sequence reports lay greater emphasis on the time element during the operational period of the event. It's a report written by merely following the sequence of time.

Further, a report may be classified into two kinds, namely technical and popular reports.

## Technical Report

A technical report is written for the fellow researcher and therefore should be organized, differently. In such a report, the researcher is expected to give a full account of the technical aspects both in the sampling methods and the subject matter. Few professionals are more concerned about the methods employed. The value of the findings and of the report it self depends on the techniques adopted. Conceptual and analytical framework, sample design should be adequately explained. M.H.Gopal mentions some distinguishing facts of a technical report.

1. The problem and its nature.
2. The method and scientific accuracy.
3. The data and their dependability.
4. The logical relationships of the data and interpretation.

These facts have to be presented logically and sequentially with simplicity, clarity and brevity.

## Planning and Organisation of a report

Before starting the report writing a good deal of planning and organising the study materials form the basis. While commencing the report writing the researcher does not know how to start writing. At the stage of writing, the researcher would have accumulated a mass of data and information which have to be prudently and carefully used. Therefore, a well conceived planning and organising, facilitates report writing with proper emphasis on different aspects. Planning permeates the entire report to each chapter, aspects and paragraph. It is an arrangement of ideas in a logical and coherent manner within the framework of the overall structure. Proper planning and organisation has the following advantages :

1) Ideas and facts which are relevant to the study alone are included.
2) The report gets greater clarity and unity.
3) Findings and results are easily communicated to the readers.
4) It facilitates transition in passing from one idea to another.
5) Ensures unity and sequence in presentation of facts, and
6) Results in an integrated and comprehensive research report.

## Steps facilitating a good organisation

Organisation of the research report involves three essential steps. (i) full acquaintance with research notes and organising it properly (ii) adequate thought about the structure of the report, reflective thinking and (iii) formation and control overnotes.

The very process of social research consists in noting down facts and making observations. Notes taking is the core activity of social research. Accuracy and utility of the notes taken from field observation has to be put to optimum use. Methodical and systematic organisation of the notes is a pre-requisite for effective report writing. While, organising research notes the following points may have to be followed by the researcher.

1) To take down notes it is better to use cards rather than sheets of paper.
2) The cards have to be assorted and arranged topically.
3) After orgainising the notes, the researcher should consider the adequacy of the notes, if he finds that the note is inadequate, he has to gather additional information.

## Formation of Satisfactory Outline

A bare outline is a 'must' in report writing. It is like 'the skeleton' in the human body. The outline of the research study is made at two stages. One at the beginning which serves as a design of the study and second before writing the report. The outline should be elaborate so as to include all important aspects which should fin a place in the report. To make the outline elaborate it should be prepared at three stages; topical outline; paragraph outline; and sentence outline. Topical outline includes the chapters and broad aspects to be included in each chapter. It is a skeleton outline. Paragraph outline includes all major paragraphs indicating the central idea of

each paragraph. Sentence outline does not imply writing of sentences. It is merely composed of points to be included in sentences. M.H.Gopal suggests the following points to be observed while planning an outline.

1) It should be as detailed as possible and should enable continuous writing;

2) It should not be vague and should not include such value phrases as 'introduction', 'body', 'conclusion', etc. which give no direction to the report writing.

3) It should fulfil the considerations of chronology, topical unity coherence and transition; and

4) Each paragraph should contain one major idea.

## OUTLINE OF A REPORT

### I. The Preliminaries

1) Title Page,

2) Preface or foreword, acknowledgments,

3) Graphs or illustrations, tables, charts,

4) Table of contents.

### II Contents of the Report

1) Introduction

a) Objectives of the study, statement of the problem, hypotheses and definition of concepts,

b) Review of literature and research studies,

c) Time, place and materials of the survey,

d) Scope, assumptions and limitations,

e) Organisation and sampling procedures,

f) Methods, tools and techniques employed for data collection.

### 2) Analysis and presentation of results

a) Report of facts - nature, volume and dimension,

b) Statistical analysis of data,

c) Interpretation, generalisation and conclusions,

d) Summary of findings and recommendations.

### III. The Reference Material

1) Bibliography

2) Appendices - Questionnaire, statistical tables, etc.

3) Glossary of Terms

4) Index

## Conclusions and Generalisation

The primary task of the researcher is that the data presented in tabulation form should first be explained. He should avoid repeating the whole data. The statistical material presented in tables should be checked for accuracy. Besides explaining tables he should interpret data on the basis of evidence. Such interpretation is the basis of generalisation. He should prefer understatement, instead of overstatement to retain readers' confidence, while drawing conclusions, the researcher should follow certain principles.

1) It is not possible to come to opposite conclusions with same material.
2) Other persons on the staff would reach the same conclusions with the evidence shown.
3) The conclusions are consistent with one another and with other known facts.
4) They seem to make sense and of not 'why'.
5) They are not to be coloured by the writer's personal biases.
6) Their meaning is unmistakable.
7) They could be accounted for by the survey techniques used.

## Arrangement of Ideas

Arrangement of ideas and systematic presentation are aids to evoke the interest on the research. The skilful writer leads his reader step by step through a well organised analysis of problem method, data and conclusion. Headings should be specific, brief and represent the points by appropriately employing centre headings and sub-headings, underpinning, bold face type etc. Besides, there must be uniformity and consistency in the presentation of ideas, punctuation, capitalisation, abbreviations, foot notes, quotations, acknowledgment and bibliography must follow a uniform pattern.

## Writing Style

Utmost attention should be paid for writing the research report. The report is a communication between the researcher and the reader. It should be clear, easy to read and accurate in information. Statements of vagueness and generalisation will make the report incomplete. The essence of communication is clarity in presentation.

A few points to be kept in mind at the time of writing a report are :

The first step clear presentation and clear thinking. Poor writing is largely due to hazy thinking. Research report writing is a reasoning process and without clarity of thought, there will not be clarity in writing.

The second step is that the researcher should have a good command over language. "Too frequently faulty sentence structure, illogical paragraphing, poor punctuation, incorrect spelling and other variations of English usage render an otherwise promising manuscript virtually worthless". At the same time, one should not use complex high sounding language at the cost of clarity. The

researcher should give importance more to clarity than style, facts than drama and incisiveness to persuasiveness.

The third point is brevity. One should take care of the economy of words and concentration of ideas. Repetition and superfluousness will abstract the lucidity of the report and will interfere with precision and clarity.

The fourth point is objectivity in report writing. One should write an unbiased, objective report fortified by facts.

Another important aspect of style is that colloquial or conversational modesty in writing are inappropriate in a research report. As Anderson and others observe : "As a general rule, personal pronouns such as I, We, You, Me,My, Our, and Us should not appear except in quotations. A thesis should not consist of the reporting of personal experience or opinion but should be critical analysis of a problem and this is better arrived at by using a tone of scientific impersonality. "Readability is another factor which a researcher should keep in mind while writing a report and sentences should not be too complex nor cramped with jargons.

## The three drafts

The research report should normally go through three drafts. The objective of the draft is comprehensiveness and fullness of facts. In the second draft the report should be improved both in the form and the language by considerable trimming and pruning. Ideas and sentences have to be realigned to get clarity, order, logic, unity, coherence and naturalness. The researcher should avoid too much technical jargons, complex language and repetition. Presentation of research data in the form of charts and diagrams enables the data in a readable form and helps the reader to understand the more complex tabulations.

The final draft concentrates on the finish and the final touch on documentation to make the report authoritative and convincing. Footnotes and bibliography makes the citations more authoritative, add to growth and continuity of knowledge and an acknowledgment of the benefits which the researcher received from others. Writing footnotes serve the following purposes :

1) To acknowledge the facts and ideas called out from other sources.
2) The researcher's contribution is distinguished from those of others.
3) The researcher is protected from being attacked by others.
4) It provides reference to a source which may be consulted for more details.
5) The larger context of the problem is indicated to the reader.

The footnotes must be relevant and the researcher must avoid the tendency to overwhelm the reader with quotations. Over documentation is bad, but under documentation is worse. Great discrimination has to be exercised in selecting the footnotes. There are, in fact, four facts of citation which, when, how much and how to cite. One should learn the guiding principles of citation. The following are a few suggestions which the researcher should keep in mind.

1) Avoid citing authority for well known facts and statements.

2) Prefer scholarly editions to casual reports.

3) Prefer citing published sources.

4) Do not repeat in footnotes what is already stated in the text.

5) If the citation is too long or contraversial one, discuss it in the appendix, and

6) If two or more previous writers have expressed one and the same idea, cite always the earlier writer.

Finally, the researcher must maintain ethical responsibility in writing the report. The researcher might have gathered information promising to maintain confidentiality to the respondents. It is the moral responsibility of the researcher to protect his source of information. Secondly, the researcher should also ensure that data presented in the report are accurate, unbiased and without distortion. Any misrepresentation or incomplete presentation, apart from damaging the development of scientific knowledge, may also bring discredit to the researcher and to the community of researchers.

In the preparation of scientific report, intellectual clarity and readability are important factors. Ultimate purpose of a report is to communicate the contents to the prospective users. Unless clarity is ensured the policy makers, practitioners, intellectuals and public in general cannot make use of it. The report writer should keep this in mind in finalising the report.

## Problems in the preparation of report

To the ordinary person, report writing is not a difficult job. It appears that it only involves gathering of facts and writing down the report in a serial order. However, the actual writing of report is not so easy. It may face several problems, of which the most important are as follows :

### 1. Problem of language

The most important problem concerning report writing is the problem of language. If the report adopts simple and local dialect, its level is considered low. Technical matters cannot be easily communicated through simple language. If too many technical and scientific terms are used in the report, the level of language of the readers may not be very high to understand the language of the report. Therefore, it is difficult to choose a language which may be simple and easy and yet convey the technical details properly. The solution of the problem of the language of the report requires a lot of patience and skill.

### 2. Problem of technical terms

Most of the physical sciences have precise and clearly defined terms. On the other hand, social sciences neither have a sufficient number of technical terms nor all the terms used in them have been precisely and clearly defined. Therefore, report writing has to face the real problem of lack of technical terms and lack of definition of the terms used.

## 3. Problems of Objectivity

Objectivity is the basic element in science. Preparation of report is a scientific job. Therefore, it should be sufficiently objective. It is often difficult to maintain sufficient objectivity due to the nature of social surveys. The physical sciences involve material subject matter. In social survey, however, the surveyor himself is a member of the subject of study. He has some preconceptions and prejudices about social matters. Therefore, it is difficult for him to be absolutely objective. He may be emotionally attracted to certain persons or groups so that he cannot objectively evaluate their merits and demerits.

## 4. The problem of publication of facts and quasi-truths

Some of the facts gathered by the researcher may be such which are true but their publication may be harmful to the interests or prestige of a particular person or group. In such a situation, the reporter has to face the dilemma to publish or not to publish the facts. The publication of quasi-truths is even more difficult. These truths have not been completely verified. Their inclusion may reduce the validity of the report, while their negation may take the report inadequate for the guidance of future investigators.

# DOCUMENTATION AND BIBLIOGRAPHY

## Meaning and Purpose

An important stage in report writing is documentation. In library science, documentation refers to the retrieval of the information pertaining to a subject, searching for the literature, preparing the list and making it available to the researcher. The work of documentation is done by librarians to enable the investigator to study the literature in order to decide whether the information is valuable for his own research. In research reporting the word documentation has a different meaning. It indicates the references to the sources, other previous and current work and views, additional data and discussion and suggested further reading on the specific problem as handled by the researcher.

Documentation serves the following purposes is a research report.

(i) It supports the facts presented or ascertains made by the investigator. Therefore, the citation must be authoritative.

(ii) It enables the investigator to acknowledge his debt to others. He should not claim as his own what is somebody else's. This also avoids legal complications which may arise due to infringement of another's copyright.

(iii) Documentation is useful and necessary to point out differences of views, facts and findings.

(iv) Documentation also leads the reader to further information and

(v) A final purpose of documentation is to provide additional or detailed information either in the footnotes or in the appendices to the report.

## Form of Documentation

Documentation has the following six components :

(A) Footnotes

(B) Bibliography

(C) Quotation and translation

(D) Tables, charts and diagrams

(E) Appendices

(F) Preface

## (A) Footnotes

Footnotes are the most important component in research documentation. As research results in growth and continuity of knowledge, the investigator benefits from other writings. To give weight and support to his own ideas or to controvert current ones footnotes are important. They serve the following needs.

(a) To acknowledge facts and ideas borrowed.

(b) To provide a reference to the sources which may be consulted for more details.

(c) To illustrate statements relevant but not important enough to be in the text.

(d) To clarify an apparent discrepancy.

(e) To give cross reference to some other part of the report.

(f) identify a quotation or citation.

(g) To report on tests of significance

(h) To explain more fully headings such as in tables.

(i) To distinguish one's own contribution from that of another and to protect one's self from attack and

(j) To acquaint the reader with the larger context of the problem.

The purpose of footnotes is definitely not to lengthen the work or to overwhelm the reader; it is essentially and completely to clarify, support or illustrate the text.

## What to cite

In selecting the footnotes the following are the guiding principles :

(i) Avoid citing authority for well-known facts or statements.

(ii) Prefer scholarly editions to casual reprints.

(iii) Prefer citing published sources.

(iv) If two or more previous writers have expressed one and the same ideas cite always the earliest writer.

(v) Do not repeat in the footnote what is already in the text.

(vi) If the citation is too long or controversial, regulate and discuss it in appendix.

### When to Cite

(1) Insert the footnote even at the first draft stage when the researcher is fresh with the facts also when citations would help him to check up his own statements.

(2) All quotations require a citation.

(3) Important fact and idea must be acknowledged fully and precisely, both for verification and for identification.

(4) Citations are necessary when there is an important exception to the material quoted or a difference of opinion.

(5) Footnotes are required when there is additional information, interesting but perhaps not completely relevant to or in the main argument.

### How much to cite

A beginner in research has a weakness to cite too much. Avoid too many and also too few footnotes. While over documentation is bad under documentation is worse. A new or controversial topic requires relatively more citations. While reporting to the fellow professional cite only on new or disputed points; while writing for the layman, pay greater attention to earlier and technically known sources. Although a citation is made, the researcher should not camouflage the ideas as his own. Make clear what is your own and what belongs to others. Make documentation, brief, clear and factual. Mention the source correctly and fully.

## (B) Bibliography

Bibliography is an essential part of a research report. This contains information about books, reports, articles and other sourfces relevant to the investigation.

A bibliography is of two kinds.

(i) The working bibliography and

(ii) The final bibliography

## Working Bibliography

Working Bibliography is formed at the initial stage of research even before the problem is selected. It is the pilot of research, its purpose being to guide the researcher in his investigation. It helps him to understand the problem, to formulate the hypothesis and to start the investigation. The working bibliography expands as research progresses and resembles, the growing collection of data.

To aid the preparation of working bibliography the following international and national bibliographies are available.

### (1) International Bibliography of Social Sciences

It is an annual publication since 1952 and is in four parts namely Sociology, Political Science, Economics and Anthropology by UNESCO in collaboration with the International Committee for Social Science Documentation.

### (2) Asia Social Science Bibliography

It is a continuing project of the Institute of Economic Growth, New Delhi since 1967 and covers education and communication, Political Science, Economics, Sociology, Social Anathropology, Social Psycology and Demography. It is a comprehensive bibliography with annotations and abstracts drawn from nearly 300 journals from Asian countries.

### (3) Documentation on Asia

It is a bibliography of articles and documentation indicative of political, economic and social developments and international relations of countries in Asia.

### (4) Indian National Bibliography

It is an authoritative bibliographical record of current publications in major languages of India. It is divided in two parts : general publications and official publications.

### (5) Bibliography of Bibliographies of India

It is a comprehensive volume comprising bibliographies, library catalogues, documentation lists, abstracts, dissertation lists etc. It has subjectwise arrangement which provides bibliographic details about each title.

## Final Bibliography

Final bibliography is prepared on the completion of the report. It serves two purposes. It helps the interested reader to pursue the problem further and enable verification of the facts and statements in the report. It also gives weight to the report by indicating the depth and extensiveness of the work done. Final Bibliography may take the following forms :

(1) A Simple list

(2) Select bibliography

(3) Classified references

(4) Critical bibliography

## (1) A Simple list

The simplest form of final bibliography is to have a simple list for books, articles reports etc. Since a list is not of much help and in fact, has been called a parade in print; for it may have the appearance but not the substance of learning.

## (2) Select Bibliography

Sometimes the researcher may select a few relevant references only and may cite them in the form of select bibliography. Such a bibliography has some drawbacks. For example, selectivity may be misleading. What the researcher discards may really be valuable to the reader working at the problem from another angle. There is also a danger of listing titles not utilised in the investigation.

## (3) Classified Reference

A third form of the final bibliography is to classify the references. The titles may be arranged according to the topics or the type of the material and may be broken up into its component sources. The break up may be of four kinds. (i) primary, secondary and territory sources (ii) printed and manuscript (iii) published and unpublished and (iv) books, reports, articles, pamphlets etc.

## (4) Critical Bibliography

Critical Bibliography is characterised by the brief, precise and suggestive comments of the researcher on each title. A well done critical bibliography brings the author and the reader close.

All types of bibliography must be adequate in volume, relevant to the problem, complete in details and informative to the reader. The length of the size of the bibliography cannot be prescribed. It depends on the nature of the problem.

The reference to any title and in the footnote and in the bibliography should be the same. In presenting references under bibliography, the following should be kept in mind.

(a) Make the list alphabetical.

(b) Put the surname of the author first and then his personal and other names.

(c) Capitalise all important words in the title and italicize the full title.

(d) For articles, capitalise the name of the journal and not the title of the article which must be within inverted commas, Invariably mention the pages in the journal.

(e) If the author is not known, put it under the first important word of title.

(g) Bibliography may be included at the end of each chapter or at the end of the completed report.

The bibliography, whatever form it takes, should, no doubt be highly selective and full but relevant. It is essential that the investigator must have read the literature and it must have some bearing on the problem.

## (c) Tables, Charts and Graphs

Tables, charts and graphs are useful especially when an investigation is statistics based. They are the most comprehensible form of presenting sizeable numerical data. Tabular presentation facilitates comparisons, which enable better appreciation of their significance. Tables, graphs and charts make the report more readable and attractive. Because of these advantages some are tempted to over do them and include a large number of long and complicated tables and graphs. But one should not forget that the report should be primarily a written language and at best these devices are only supporting, explanatory and secondary in purpose.

He should therefore bear in mind the following :

(1) Be selective and discriminating in introducing tables and other aids in the final report.

(2) Make each table self-explanatory.

(3) Better limit each table to a singly page.

(4) Do not repeat in the table what is found in the text.

(5) Space tables to facilitate reading them and also the text.

(6) Avoid statistical sophistication.

(7) If a large number of tables is necessary simplify each to make it clear. Weigh the gain in clarity against the loss in detail.

## (D) Quotations

Quotations from other works, common , and often necessary in research reports serve two purposes; (1) Facts may be taken over from another source, either because it is fuller or because the facts are relevant to the current investigation but not as its major substance. (2) Opinion or facts may be cited either to support the investigator's own findings or to be controverted by him. Quotations take one or two forms : (1) in the original language or (2) in translation with or without the original passage. Too many and too long quotations suggest the researcher's evasion of responsibility and inability to do him own thinking. The investigator who quotes too much, ceases to be a writer and becomes a compiler. He must not forget that the report should be his own work. Therefore, the quotations should be few, short and directly relevant to the point under consideration.

## (E) Appendices

Appendices are useful for placing cumbersome material, breaking the continuity of the main text. They are not the place for dumping irrelevant material or for making the report bulky. They are usually at the end of the text, but occasionally, also at the end of a part.

The appendices are either technical or non-technical. Technical appendices include materials relating to methodological discussions of techniques, elaboration of concepts, term and definitions and mathematical derivations and formulae. The non-technical appendices include information useful and relevant but too elaborate for the text, supporting data and lengthy tables, lengthy quotations and forms of questionnaires and schedules.

## (F) Preface

A preface is not essential to research writing, but by convention it has become a part of the report. The preface gives an opportunity to the writer to acknowledge his indebtedness to his helpers of various types and for different senses. An investigator is of course, indebted to many persons in different degrees and it is his pleasant duty to thank them. But a weakness noticeable in many Indian publications is the elaboration of such acknowledgments and the different degrees of obligations expressed. Even if there are varying degrees of sincerity of thanks and of gratitude, the acknowledgments could be simple and yet no less sincere.

## Model Questions

1. Explain the various components of documentation.
2. Write a essay on the structure and writing of a research report.

## BOOKS REFERRED


1. Ram Nath Sharma : Techniques of Social Research, Rajhans Prakashan Mandir, Meerut.
2. R.N.Sharma and R.K.Sharma : Reseach Methods in Social Sciences, Media Promoters and Publishers Pvt. Ltd., Bombay, 1983.
3. M.H.Gopal, An introduction to Research Procedure in Social Sciences, Asia Publishing House, Bombay, 1970.R
4. Margaret R.Iverson : The Research Paper Simplified, Prentice Hall of India Private Limited, New Delhi, 1970.
5. C.R.Kothari : Research Methodology Methods and Techniques, Wiley Eastern Limited, New Delhi.
6. B.C.Tandon : Research Methadology in Social Sciences, Chaitanya Publishing House, Allahabad, 1979.

## LESSON - 8

# METHODS OF PRESENTATION OF STATISTICAL INFORMATION - I

## DIAGRAMS

## NEED, USEFULNESS AND LIMITATIONS OF DIAGRAMMATIC AND GRAPHIC REPRESENTATION OF DATA

A chart (i.e.,) a graph or diagram is the most forceful medium of presenting statistical data. Evidence of this can be found in the financial pages of newspapers and political journals, advertisements etc. The special feature of diagrams and graphs is that they do away with figures altogether and present the dry and uninteresting facts in the shape of attractive and appealing pictures. Diagram gives the reader a clear and simple impression which he cannot get from reading the same material in a table or text. Hence, diagrammatic and graphic representation of data is gradually becoming popular. These methods were first introduced by William Playfair in 1786.

### Usefulness of Diagrams and Graphs

1. Diagrams are attractive and create lasting impression. This is human psycology and a successful advertiser or propagandist always exploits this psychology of the people to win his mark.

2. Diagrams and graphs render comparison easy as a man can easily compare lengths, areas and volumes. It would help the people considerably if the results are presented diagrammatically. At a glance one can have an idea of the progress. If adequate charts are maintained, the charts will be readily intelligible at a glance. Also, it will not only help to deal with the cases more efficiently but in saving a considerable amount of his valuable time and energy.

3. Diagrams will be of much utility to those who do not want to go into detail for want of time but still wish to have some idea.

4. Diagrams are very useful in all types of studies. A purely theoretical economist may easily understand the basis for logical reasoning and easily explain as economic law with the help of graphs. A practical economist may make his ideas impressive through diagrammatic representation. Social reformers, politicians administrators and infact all types of people greatly benefit by the use of diagrams and pictures in their respective fields.

5. With the help of graphic curves, it is possible to forecast tendencies and likely changes in the future based on the experience in the past. Looking to the growth of the population in the past, one could estimate the population figures for the future.

6. Suppose we have a set of data for a series of years, except for one particular intermediate year. The missing figure for that intermediate year can be more easily found out with a fair degree of accuracy with the help of curves.

7. Inequalities in the distribution of income and wealth, correlation between variables, trend, seasonal variations and cyclical variations in time series etc., can be easily marked by graphic diagrams.

8. Graphs help us in locating statistical measures, such as median, mode, quartiles etc.

9. In statistical quality control, charts are very much useful in separating chance variations from assignable variations.

## Limitations of diagrams and graphs

1. Diagrams can only supplement but cannot substitute tables.

2. Diagrams or graphs cannot show so many sets of facts as may be shown in a table. Numerous rows and columns may appear in a table. But a diagram show only a limited amount of information.

3. Although exact values can be given in a table, only approximate value can ordinarily be shown by a diagram. In a table we may enter as many digits as are desired but we can plot only the approximate value on a diagram.

4. Diagrams are useful for giving a quick picture of a general situation, but not details. It is very difficult to note the difference between Rs.289 crores and Rs.296 crores of export when both are represented by small diagrams.

5. If the difference between measurements is very much say, 10 : 10,000, the comparative presentation by diagrams is impossible.

6. It is not unusual to find the same data being represented differently by different parties having opposite interests and thus presenting quite different pictures. With the change of scales, a graph or diagram will readily admit diminution or exaggeration of facts.

7. Graphs and diagrams take more time to construct than a table since each diagram is an original drawing. This difficulty however is offset by the added effectiveness which the diagram possesses in comparison with table.

8. The main difficulty with diagrams and graphs is that many people are not accustomed to them and they do not attach much importance to them. Pictorial, diagrammatic and graphic forms are not taken as a mass of data properly tabulated.

9. Diagrams can be as much misused as they are useful because in advertisement and political propaganda they are often based on motive. Hence, the true statistician has to guard himself against misinterpretation.

# GENERAL RULES TO BE FOLLOWED IN DRAWING DIAGRAMS AND GRAPHS

There are several types of diagrams and graphs and the technique of drawing them will differ. However, there are certain general directions. They are as follows.

## Planning the chart

One of the important requirements is to decide upon the purpose of the diagram. Other requirements are the type of diagram, its size, proportions, form and the preliminary if not final title.

A diagram should achieve simplicity and the reader's point of view must be kept in mind. What is simple for one reader may be complex for another. In case of doubt, it is wiser to oversimplify a chart than to make it unduly complex. Also technical terms should be avoided as far as possible.

## Using ruled Paper

Diagrams are generally drawn on plain paper. But, to draw graphs specially ruled paper known as graph paper or coordinate paper is used. The graph paper will contain horizontal and vertical lines. Before drawing a curve, one of the horizontal lines and one of the vertical lines are picked up and named as X -axis and Y - axis respectively. The point where the X - axis and Y - axis cut each other is named the Origin and denoted by the letter O. Positive values and shown to the right of O on the X - axis and above O one the Y - axis. Negative values are shown to the left of O on the X - axis and below O on the Y - axis. The positive and negative values on the axis increase as we move away from the origin. The exact position of any point can be located by measuring from the X - axis and from the Y -axis.

In the case of graphs depicting chronologically classified ( ie. time series) data, time units ( for example, years or months) are marked on the X -axis and the amount on the Y - axis. On the other hand in the case of a graph of frequency distribution, size of each class or class limits are shown on the X - axis and the number in each class (ie.) Frequency is marked on the Y -axis. This type of marking is based on the convention that the independent variable is marked on the X - axis and the dependent variable is marked on the Y -axis. That is, the horizontal scale defines the question and the vertical scale shows the answer.

## Scale

To draw both diagrams and graphs a convenient scale must be selected. In the case of diagrams, only one scale is required to show different sizes. Scale must be clearly indicated in the right hand top corner or below the diagram. For example, the scale may be indicated as 1 cm. represents 'one thousand number of person's or 1 sq.cm represents 'one lakh units of output'.

In the case of graphs two scales - one for the X -axis and the other for the Y -axis are to be chosen. X -scale and Y - scale need not be the same. The scale on Y - axis generally starts from zero whereas on X -axis it starts with the first time period or the first lower class limit. As in the case of diagram, here also the scales selected must be noted at the right hand top corner or below the graph.

## Size

The size of a diagram depends on the amount of details involved and the importance of the diagram. When a diagram includes a great deal of information, it must be large enough for easy reading.

## Proportion

Proper proportions in a diagram is more important than size because distortion can spoil both the appearance and the meaning of a finished diagram. If the diagram is too long and narrow, either vertically or horizontally, it will have an awkward, stretched out appearance and will unduly exaggerate one aspect of the data. Hence, the diagram must be well proportioned.

## Lettering

There is no part of the diagram more important than the lettering. All lettering should be placed horizontally, if possible. Occasionally space limitations may necessitate placing the vertical scale or Y -scale label in a vertical position, but such a limitation is not often present. All lettering should be legible. Lettering should not be so heavy as to detract from the diagram itself.

## Title

The title is also more important in a diagram if it is prepared for publication. The title should give the subject of the diagram clearly and completely, covering any question that may arise. The title may be given either at the top of the diagram or below it.

## Accuracy

Accuracy is graphs is more a matter of portraying a true picture to the reader that showing exact values. Hence, care must be exercised to avoid unconscious misinterpretation.

## General Appearance

Confidence in a chart is much greater if it is neat and accurate. The diagram must be drawn with the aid of geometrical instruments. Different types of lines, shades, colours, etc. can be used to make the diagram more attractive. A border line may be drawn around the graph to give it a finished appearance.

## Explanatory material

When different types of line, shades, colours etc. are used in a diagram an index illustrating those should be given for easily making out the meaning of the diagram.

If possible, the original data should be given on the same page of the diagram or on the page facing it. This will help the observer to see the details with accuracy.

Each diagram should contain a source reference to indicate the author, title, volume, page, publisher and date of the publication from which the data were taken.

# BAR DIAGRAMS AND AREA DIAGRAMS

Several types of diagrams are used for the representation of statistical data. The following are some of the important types of diagrams.

1. One dimensional diagrams or bar diagrams.
2. Two dimensional diagrams or rectangles, squares and circles.
3. Pictures or pictograms.

## Bar Diagrams

A basic form of diagrammatic representation is bar diagram. The term bar, is used to denote a thick wide line. Only the length of the bar is used to represent the given magnitude. The width of the bar is shown merely to make the diagram look more beautiful and explanatory.

Bar diagram may be used to represent any type of data. But mostly they are used to represent only geographically and qualitatively classified data.

## Method of construction of bar diagrams

1. Bar diagrams may be drawn on plain paper but it is better on a graph paper. Because, it will facilitate accurate determination of the length of the bar.
2. All the bars should be drawn with the same width.
3. The length of each bar is taken in such a way that, it is proportional to the magnitude which it represents. For example, suppose 10,000 and 1,000 are two magnitudes to be represented. If the first magnitude viz. 10,000 is represented by a bar or 10 cms length then the second magnitude must be represented by a bar of 1 cm. length.
4. Proper scale must be selected. The appearance of the diagram depends entirely on the scale. The scale should be adjusted in proportion to the magnitude of the biggest item and size of the diagram.

   Scales should be laid out in round numbers of convenient intervals such as multiples of 5, 10, 25,100, 1000 etc. The units represented by the scale such as 'crores of rupees', 'number of students' should be indicated clearly along the scale.
5. The scale should always start from zero in order to present a true comparison.
6. Equal spaces should be left between bars and at both ends (ie.) before the first bar and after the last bar. The gap between the bars should not be less than half the width of a bar and not greater than the width of a bar.
7. All the bars should rest on the same line called the base line.
8. The bars may be drawn either vertically or horizontally. However, time series data are always represented by vertical bars while geographically and qualitatively classified data are mostly represented by horizontal bars and sometimes by vertical bars.

9. When vertical bars are drawn, the scale is given at the left hand side of diagram. If there are too many bars, then the scale is given at both the sides. When horizontal bars are drawn, the scale may be given either above or below the bars.

10. The bars should be arranged from left to right (or from top to bottom) in descending order of magnitude. However, in the case of time series data, bars are to be given only in the chronological order and not according to magnitude.

In the case of bar diagram only the length of each bar (and not the width) is taken into account. Hence they are known as one dimensional diagrams.

There are various types of bar diagrams. They are :

(1) Simple bar diagram

(2) Multiple bar diagram

(3) Sub-divided or component bar diagram

(4) Percentage bar diagram.

## Simple bar diagram

This is used to represent different sizes having no sub-divisions. That is the length of each bar represents a whole magnitude. For example, data relating to Government expenditure under different heads in a particular year or sales figures of a company for various years, or production of rice in different states of population figures for various countries may be shown by means of simple bar diagram.

In the case of simple bar diagram, the bars may be left blank. However, bars are most effective when shaded to stand out against the white background.

## Illustration - I

The following data represent investments in the Second Five Year Plan in the Public Sector. Draw a suitable diagram.

| Items | Investment in crores of Rs. |
|---|---|
| Agriculture | 568 |
| Irrigation | 913 |
| Industries | 890 |
| Transport | 1385 |
| Social Services | 945 |
| Miscellaneous | 99 |

The above data can be suitably presented by simple bar diagram, the length of each bar representing the amount of investment under each head.

As the given data are qualitatively classified data, let us first rearrange the items in descending order of magnitude as follows :

| Items | Investment in crores of Rs. |
|---|---|
| Transport | 1385 |
| Social Services | 945 |
| Irrigation and Power | 913 |
| Industry and Mining | 890 |
| Agriculture | 563 |
| Miscellaneous | 99 |

Let us draw the bars horizontally marking the items of investment on the Y -axis and the amount of investment on the X -axis.

Let us take the scale, 1 cm = Rs.289 crs.

**Fig. 1**

**Investment in the Second Five Year Plan in the public sector**

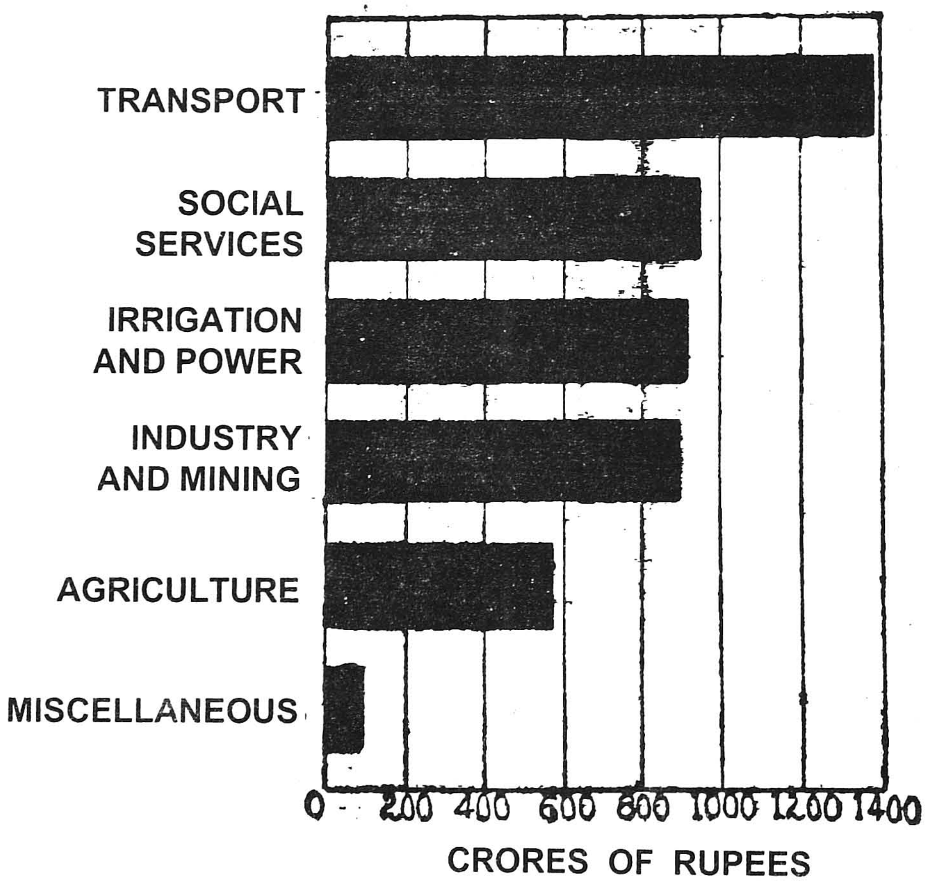


**Illustration - 2**

Represent the following data by means of bar diagram.

Sales of A firm from 1965 to 1968

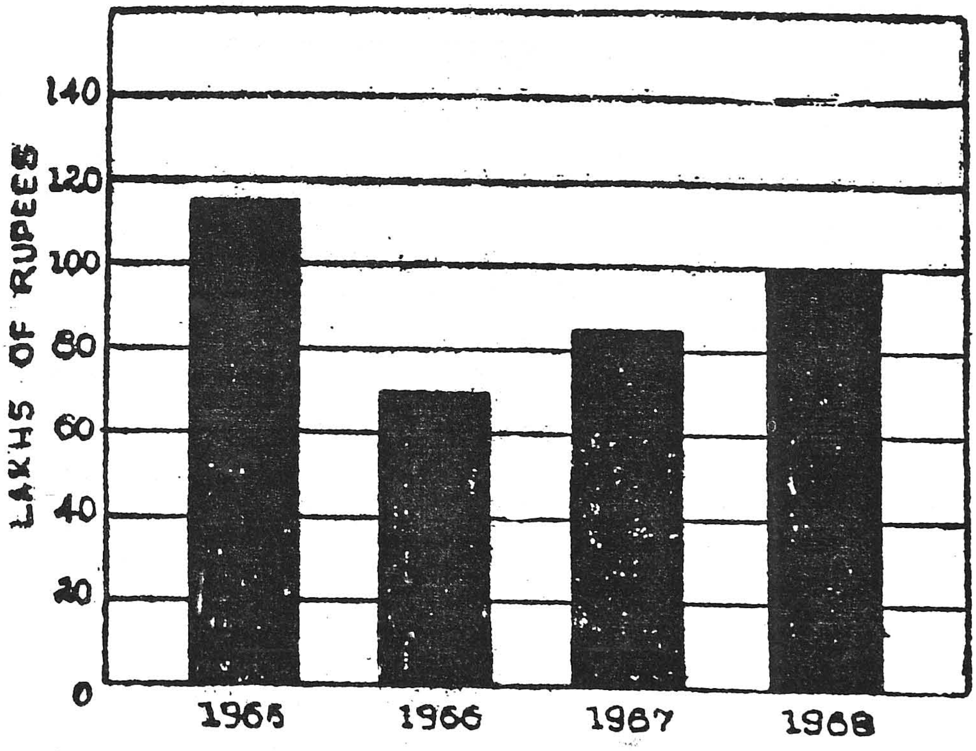
| Year | Sales (in lakhs of rupees) |
|---|---|
| 1965 | 115 |
| 1966 | 70 |
| 1967 | 85 |
| 1968 | 100 |

Note that we are given time series data. So we should not rearrange the data according to order of magnitude. Also, we should represent the data only by vertical bars and not by horizontal bars.

Let us take the scale, 1 cm = Rs. 20 lakhs.

**Fig. 2**

**SALES OF A FIRM - 1965 to 1968**




## Line diagram

If the number of items is very large, bars can be replaced by lines. But such line diagrams lack popularity, because the lines have appreciable width and their vision is hence limited to a great extent.

An important limitation of simple bar diagram is that it can show only one classification of data. If the given data consist of two or more sub classifications these cannot be shown in simple bar diagrams. If we are given total population with sex-wise distribution ie. total number of males and total number of females for various decades, these data cannot be shown by simple bar diagram.

## Multiple bar diagrams

When the given data consist of two or more sub-classifications, they may be shown by grouping the bars. Such type of diagrams are known as multiple or compound bar diagrams.

Suppose we are given quarterly production of coal ( in 1000s of tons) for a period of five years from 1954 to 1958. Here for each year, we can draw a group of four bars represent the production of coal into the I, II, III and IV quarters respectively.

Each bar in a group is distinguished by a different colour or by a different shading. However, the same colour or the same shading must be used for the first bar in all the groups. The same is true of the second, third and fourth bars in all the groups. An index illustrating the different colours or shade used is given below or by the side of the diagram.

Equal gap is left between any two groups of bars but no gap is left between bars within a group.

All the bars are drawn on a common scale so that their lengths would be comparable.

Multiple bar diagrams enable us to compare more than one set of comparisons. In the above example, each year, we can compare the production of coal in one quarter with the production figures for the remaining three quarters. In addition, the production of coal in each quarter can be compared from year to year.

## Illustration - 3

We have given below the Table representing the quarterly productions of coal (in 1000s of tons) for a period of five years.

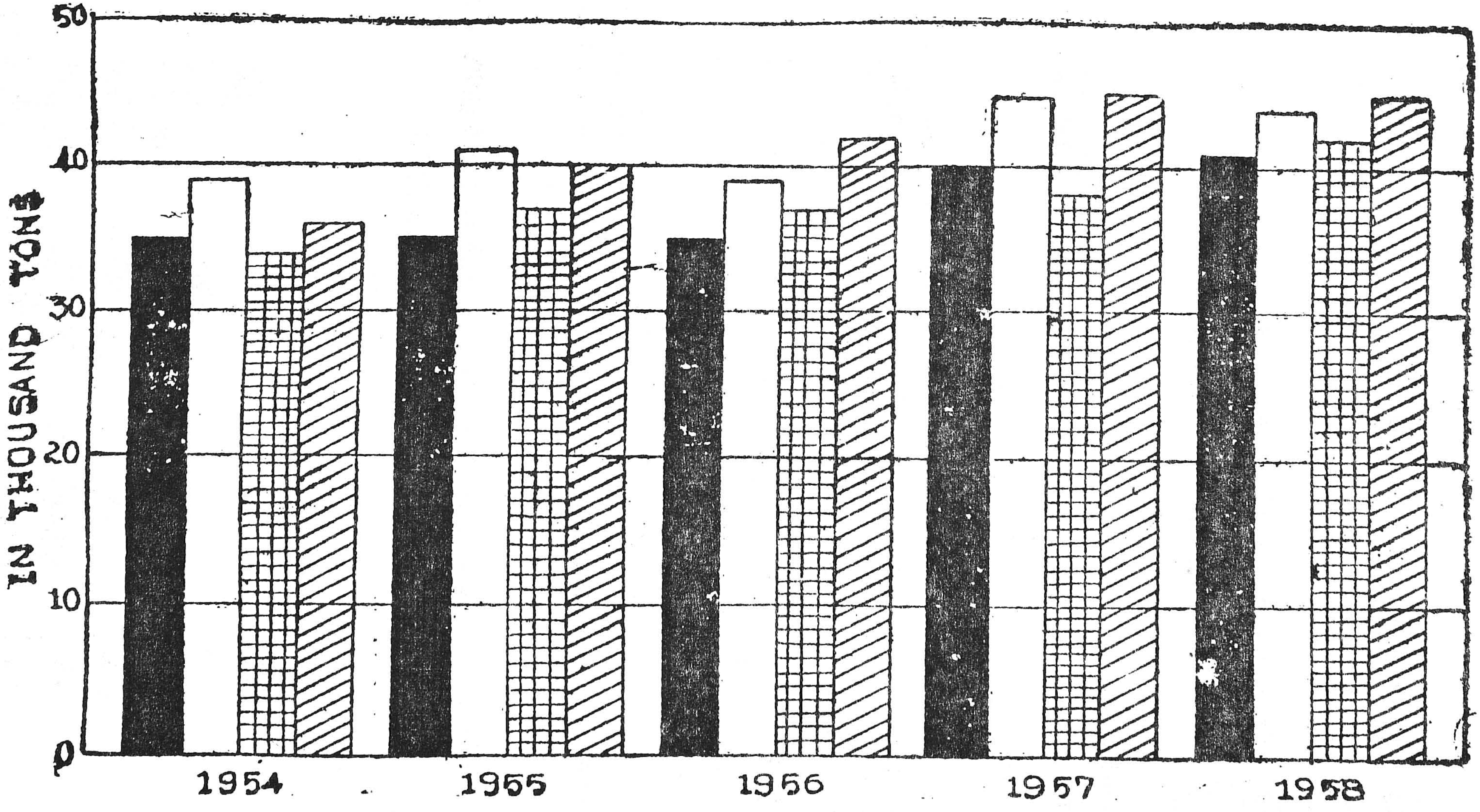
| Year | I | II | III | IV |
|---|---|---|---|---|
| 1954 | 35 | 39 | 34 | 36 |
| 1955 | 35 | 41 | 37 | 40 |
| 1956 | 35 | 39 | 37 | 42 |
| 1957 | 40 | 46 | 38 | 45 |
| 1958 | 41 | 44 | 42 | 45 |

We can represent the above data by multiple bar diagrams as follows :

Fig. 3

QUARTERLY PRODUCTION OF COAL FROM 1954 to 1958

Scale : 1 cm = 10 thousand tons



**Index**

Production of coal in I quarter

Production of coal in I quarter II quarter

Production of coal in I quarter III quarter

Production of coal in I quarter IV quarter

## Illustration - 4

The following Table shows the results of M.A. (Economics) students of a university for three years.

Number of students

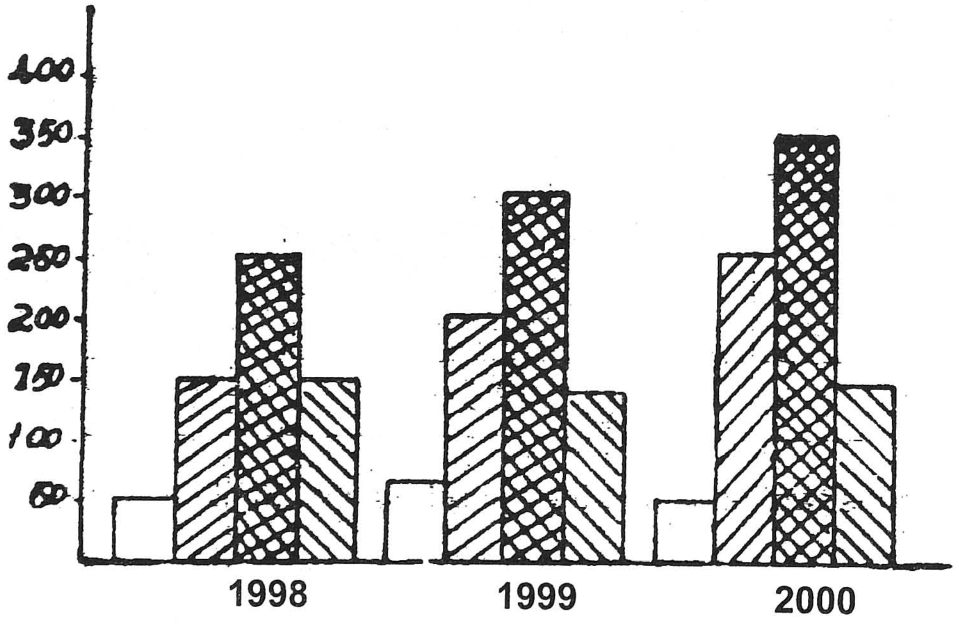
| Year | I Class | II Class | III Class | Failed |
|---|---|---|---|---|
| 1998 | 50 | 150 | 250 | 150 |
| 1999 | 60 | 200 | 300 | 140 |
| 2000 | 50 | 250 | 350 | 150 |

Represent the above data by a suitable diagram.

Given Table can be suitably represented by means of a multiple bar diagram as follows.

**Fig. 4**

**RESULT OF M.A (Economics) STUDENTS OF A UNIVERSITY FOR THREE YEARS 1998 to 2000**



## Index

I Class

II Class

III Class

Failed

## Limitation of multiple bar diagrams

In illustration - 3, suppose we are given total annual production figures also in addition to quarterly production figures. These annual totals cannot be shown in multiple bar diagrams. Thus, multiple bar diagrams cannot be used to show totals as well as their subdivisions.

## Sub-Divided or Component Bar Diagrams

Bar diagrams may be used effectively to show the component parts of a total as well as the total itself. They are known as sub-divided bar diagrams. Here, the bars are drawn in such a way that their lengths are proportional to the totals and they are divided in the ratio of their components. The order in which components are shown in a 'bar' is followed in all the bars used in the diagrams. Each subdivision in a bar is given a distinct colour or shading. The first sub-division in all the bars are given the same colour. The same is true of the second, third etc. sub-divisions also in the bars. All index illustrating the different colours or shades is given by the side of the diagram or below the diagram.

Consider the following illustration.

## Illustration - 5

The following Table shows the development expenditure (in lakh of Rs.) of part A and part B states of the Indian Union, in the year 1953-54.

| | Rupees in lakhs | |
|---|---|---|
| | Part A States | Part B States |
| Agriculture | 10,231 | 2,171 |
| Transport | 2,052 | 274 |
| Industry | 235 | 223 |
| Miscellaneous | 160 | 9 (approx.) |
| Total | 12,678 | 2,777 |

The above Table contains two totals and their subdivisions. So it can be suitably represented by subdivided bar diagram.

To draw the sub-divided bar diagram, first we draw two bars whose lengths are in the ratio of the two totals given (ie.) in the ratio 12,678 : 2,777. That is, if we take the scale 1 cm = Rs. 100 lakh. Then the first bar with 12.7 cms. (approximately) as its length, and the second bar with 2.8 cms (approximately) as its length.

Next, we mark the subdivisions inside the two bars as follows.

The value of the first subdivision of the first total is 10,231. It is marked inside the first bar by drawing a horizontal line at a height of approximately 10,. cms from the line.

2,052 is the value of the second subdivision of the first total. It is marked above the first subdivision which we have already marked inside the first bar. That is above the first subdivision we measure a height of approximately 2.1 cms and draw a horizontal line inside the first bar.

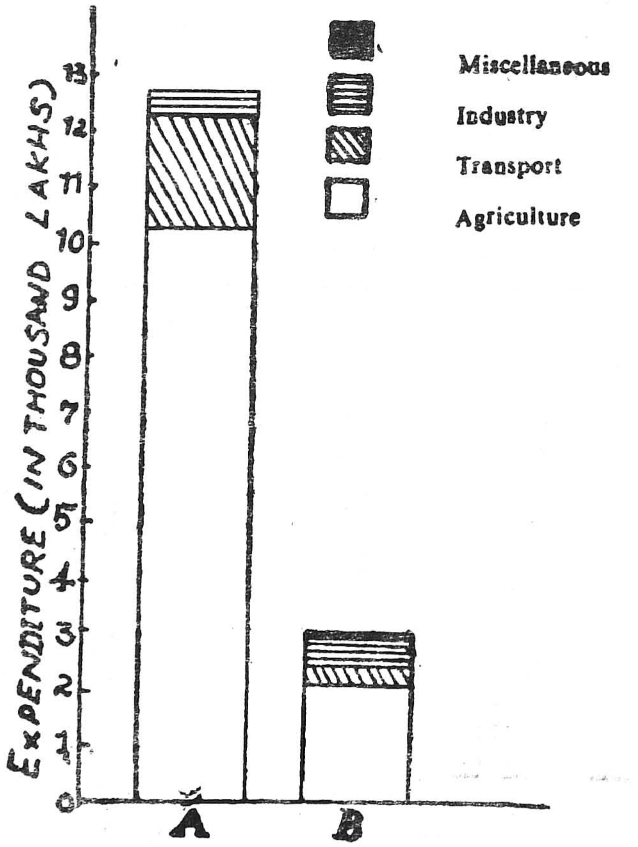
In the same way, we mark the remaining subdivisions also inside the first bar.

To mark the sub-divisions inside the second bar also, we follow a similar procedure.

In the following diagram, we have left the first sub-division blank. For the rest of the sub-divisions we have given different shading. An index illustrating the different shadings used in diagram is given by the side of the diagram.

**Fig. 5**

**DEVELOPMENT EXPENDITURE (IN LAKHS OF RUPEES PART A AND PART B STATES OF INDIAN UNION (1958-59)**




**Scale : - 1 unit Rs. 1000 lakhs**

## Illustration - 6

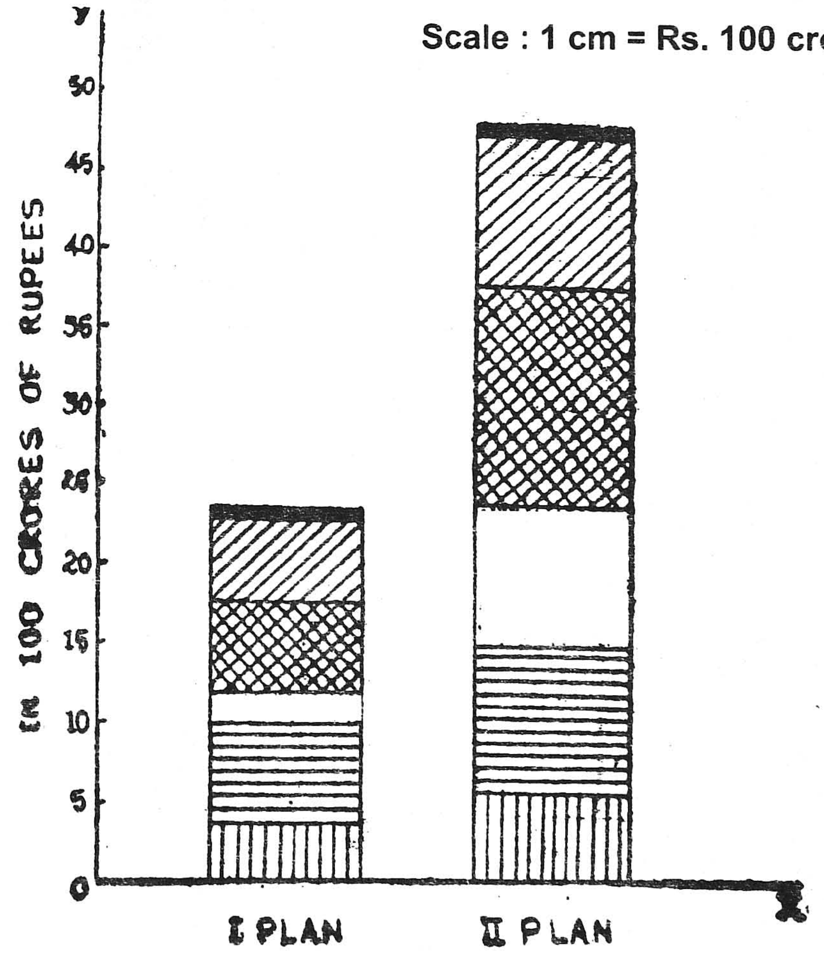
The following Table gives the distribution of outlay in the First and Second Five Year Plans of India under major heads of development expenditure. Represent these figures by a suitable diagram.

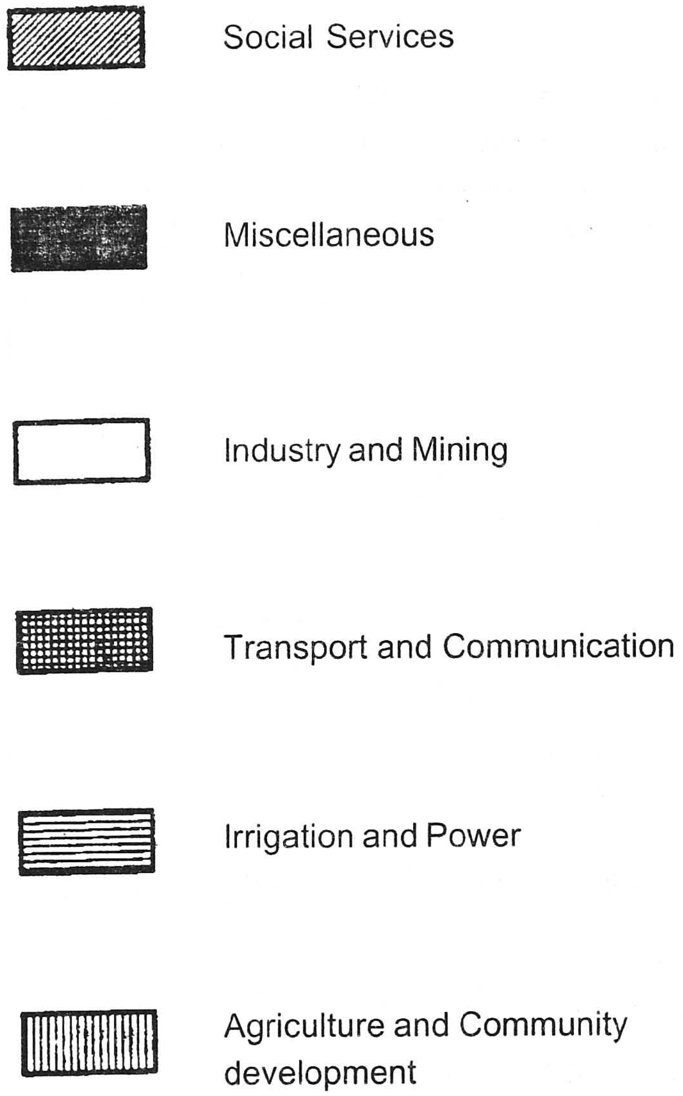
| Heads of expenditure | First Plan Rs. (crores) | Second Plan Rs. (crores) |
|---|---|---|
| Agriculture and Community Development | 357 | 568 |
| Irrigation and Power | 661 | 913 |
| Industry and Mining | 179 | 890 |
| Transport and Communication | 557 | 1385 |
| Social Services | 533 | 945 |
| Miscellaneous | 69 | 99 |
| Total | 2356 | 4800 |

The above Table can be suitably represented by a sub-divided bar diagram as follows :

**Fig : 6**

**DISTRIBUTION OF OUTLAY IN THE FIRST AND SECOND FIVE YEAR PLANS**






## Percentage bar diagram

In all the diagrams given earlier, bars have been used to represent the actual or absolute values. But, on many occasions, we have to portray relative changes in the values of components. In such cases, each total in the given data is equated to 100 and each subdivision is calculated as a percentage of the total. Now, the percentage value of components are represented by means of subdivided bar diagram and this diagram is known as 'percentage bar diagram'.

All the bars, in a percentage bar diagram will have the equal height. Because, they represent the same total, namely, 100 percent.

## Illustration - 7

Consider the same Table given in illustration - 6 and their subdivisions in absolute terms. But to draw the percentage bar diagram, we require the subdivisions in percentage terms. So first we calculated the given components as percentages of the respective totals.

Percentage value of the amount spent on Agriculture and Community Development in the First Plan

$$= \frac{357}{2356} \times 100$$

$$= 15.1\%$$

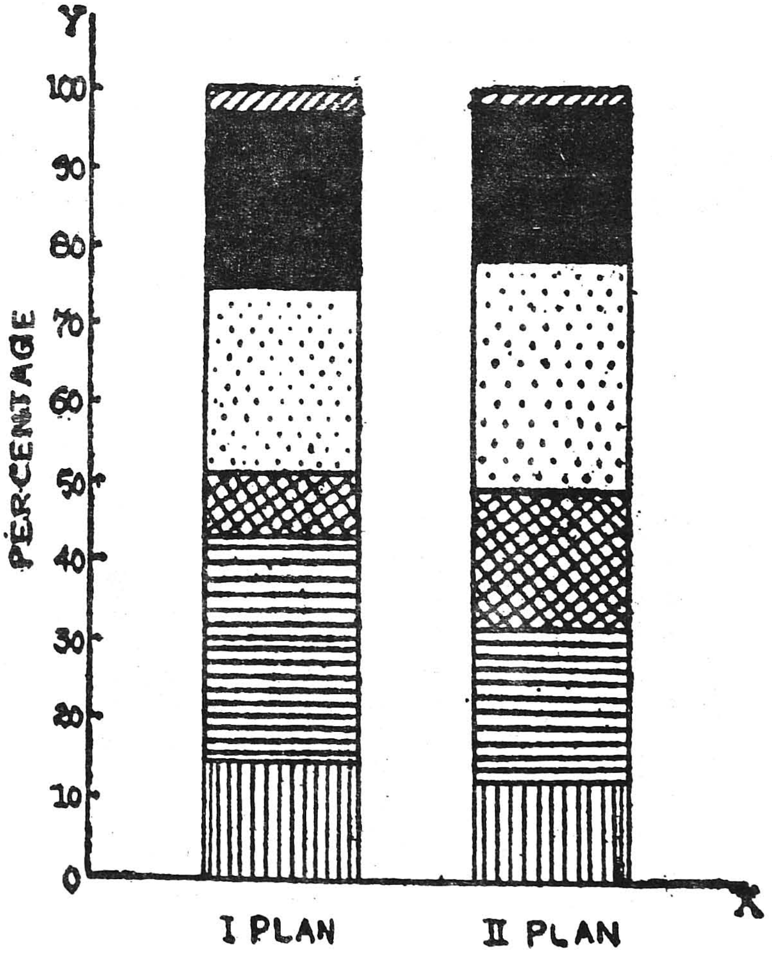
Percentage values of the remaining sub-divisions are also calculated on a similar manner and we have given them in the following table.

| | First Plan | | Second Plan | |
|---|---|---|---|---|
| Heads of Expenditure | Rs. (crores) | % Rs.2356 crs = 100 | Rs. (crores) | % Rs.480 crs = 100 |
| Agriculture and Community Devlelopment | 357 | 15.1 | 568 | 11.8 |
| Irrigation and Power | 661 | 28 | 1913 | 19.0 |
| Industry and Mining | 179 | 7.6 | 890 | 18.5 |
| Transport and Communication | 557 | 23.6 | 1385 | 28.9 |
| Social Services | 533 | 22.6 | 945 | 19.7 |
| Miscellaneous | 69 | 3.0 | 99 | 2.1 |
| Total | 2356 | 100 | 4800 | 100 |

**Fig : 7**

**DISTRIBUTION OF OUTLAY IN THE FIRST AND SECOND FIVE YEAR PLANS**

**Scale : 1 cm = 10%**



Miscellaneous

Social Services

Transport and Communication

Industry and Mining

Irrigation and Power

Agriculture and Community development

## Illustration - 8

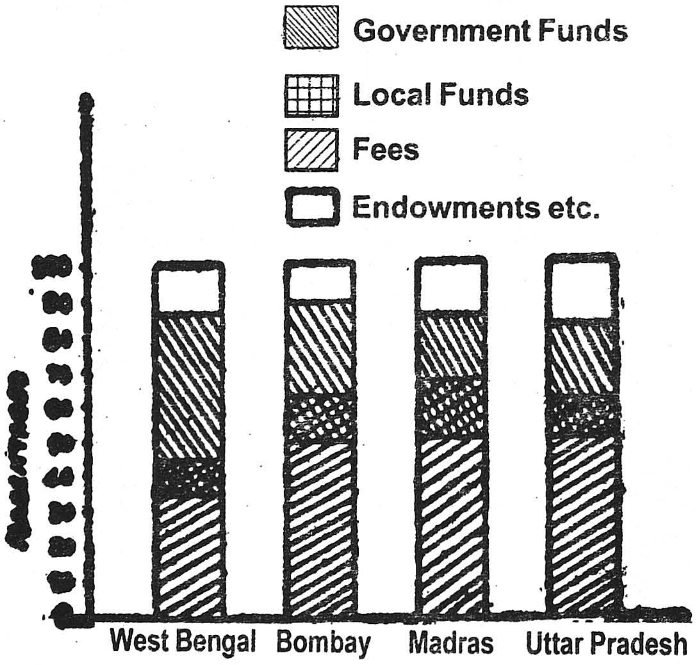
Draw a percentage bar diagram to show the expenditure on education from different sources for four important states in the Indian Union in 1947-48 given below in the form of a table.

| States | Government Funds | Local Funds | Fees | Endowments etc. | Total |
|---|---|---|---|---|---|
| West Bengal | 35 | 10 | 41 | 14 | 100 |
| Bombay | 50 | 13 | 26 | 11 | 100 |
| Madras | 1 | 16 | 18 | 15 | 100 |
| Uttar Pradesh | 46 | 11 | 26 | 17 | 100 |

**Fig : 8**

**EXPENDITURE ON EDUCATION FROM DIFFERENT SOURCES FOR FOUR IMPORTANT STATES IN THE INDIAN UNION 1947 - 48**

**Scale : 1 cm = 10%**



## Limitation of percentage bar diagram

The main defect of percentage bar diagram is that it fails to show the difference in the total values as all the totals are taken to be 100. This defect is removed by rectangular diagram which is discussed in the next lesson.

## Bilateral bar diagram

When we are given the data relating to the costs, sale proceeds and profit and loss of a company for two or more years, they can be suitably represented by percentage bar diagram. Hence, the bar or bars representing losses will have a portion below the X-axis also. As bars are used here to show both plus and minus directions, this type of percentage bar diagram is known as Bilateral bar diagram. It is also known as 'gain or loss chart'.

## Illustration - 9

Represent the following data by a suitable diagram :

| | 1950 | 1960 |
|---|---|---|
| Sale proceeds per chair | 100 | 150 |
| Cost per chair : | | |
| Wages | 30 | 90 |
| Material | 40 | 40 |
| Other costs | 20 | 30 |
| Total costs | 90 | 160 |
| Profit (+) or loss (-) | + 10 | - 10 |

We are given absolute values of cost, sale proceeds and profit and loss for two years. But, in order to make proper comparison between the costs, sale proceeds and profit and loss, only percentage bar diagram is most suited. So, first of all, we convert the given data into percentages of proceeds (ie. by equating sale proceed to 100) for each of the two years.

| | 1950 | | 1960 | |
|---|---|---|---|---|
| | Rs. | % | Rs. | % |
| Sale proceed | 100 | 100 | 150 | 100 |
| Cost per chair : | | | | |
| Wages | 30 | 30 | 90 | 600 |
| Material | 40 | 40 | 40 | 26.7 |
| Other costs | 20 | 20 | 50 | 20.0 |
| Total costs | 90 | 90 | 100 | 106.7 |
| Profit (+) or loss (-) | +10 | +10 | -10 | -6.7 |

To represent the above Table by means of bar diagram, first we draw two bars with the same height representing 100 %. If we take the scale, 1 cm = 10 % then two bars each with the same 10 cms height are drawn. Now, we mark the subdivisions as follows :

In the case of sub-divided and percentage bar diagram we have marked the components by moving to the upward direction starting from the baseline. One the contrary, in the case of bilateral bars we start marking the subdivisions from above in the downward direction.

Out of 100% sale proceeds 30 % goes to wage payment in the year 1950. This is marked inside the first bar by drawing a horizontal bar at a height of 3 cms measured from above in the downward direction.

Below this first subdivision, we measure a height of 4 cms in the downward direction and draw a horizontal line. This represents the material cost in the year 1950 namely 40%.

Below the second subdivision, we measure a height of 2 cms to denote the component namely, other costs.

Now, we are left with a portion having 1 cm height at the bottom of the first bar. This portion represents the profit which is 10% of sale proceeds in the year 1950. Here ends the marking of subdivision inside the first bar.

Now, let us pass on to the second bar. Here also, we follow the same procedure.

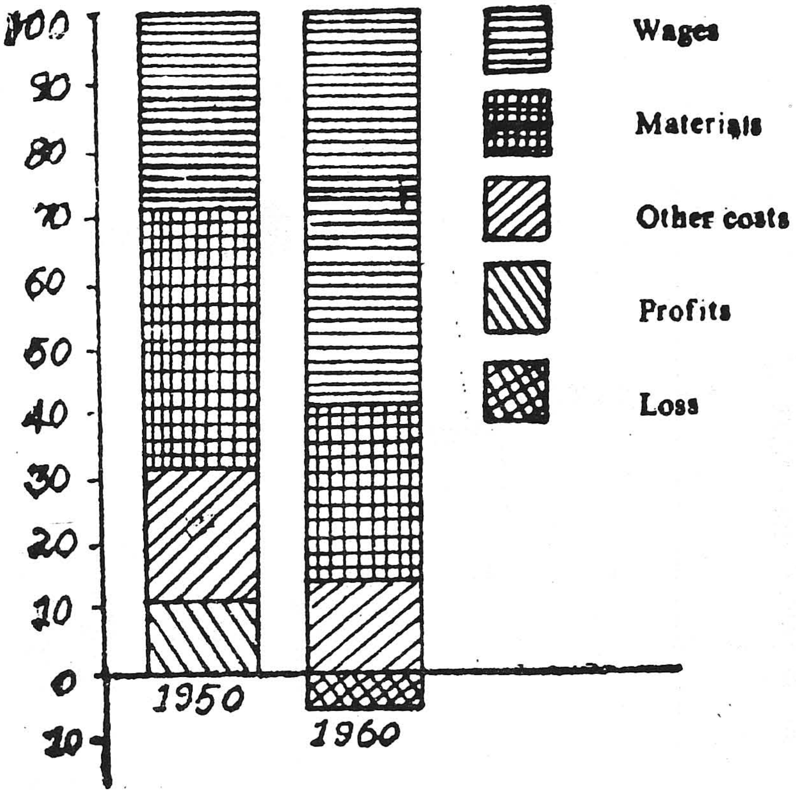
From above we measure a height of 6 cms in the downward direction and draw horizontal line inside the second bar. This denotes wage payment which constitutes 60 % of total sale proceeds in the year 1960. Below this, we measure a height of approximately 2.7 cm and draw a horizontal line. This denotes the material cost which is 26.7 % of total sale proceeds in 1960. Now we are left with only 1.3 cm (= 10 - 6 -2.7 = 10 -.78) at the bottom of the second bar. But the value of the third component namely, other costs in 1960 is 20% and to mark this component we require 2 cms height. So we extend the height of the second bar below the X-axis in such a way that the height of the bar below the second subdivision becomes 2 cms. That is, we extend the second bar below the X-axis to a height of 7 cm.

The portion of the second bar below the X-axis represents the loss, namely 6.7% in the year 1960. It is to be noted that the loss is shown as part of other costs.

## Fig : 9

**Percentage cost, sale proceeds, profit or loss per chair during 1950 and 1960**

**Scale : 1 cm = 10%**




## Deviation bars

Bar diagram may be used to represent quantities (ie. excess or deficit). For example, net profit or loss, net exports or imports etc. may be represented by means of bar diagram. Here the bars will represent both positive and negative values. Positive values, are shown above the base line and negative values are shown below the base line. As the bars represent here only the difference or deviations the diagram is called "deviation bar diagram".

## Illustration - 10

For the data given below, draw the deviation bars.

Value of Indo - U.S. Trade During 1960-65
(Rs. in Million)

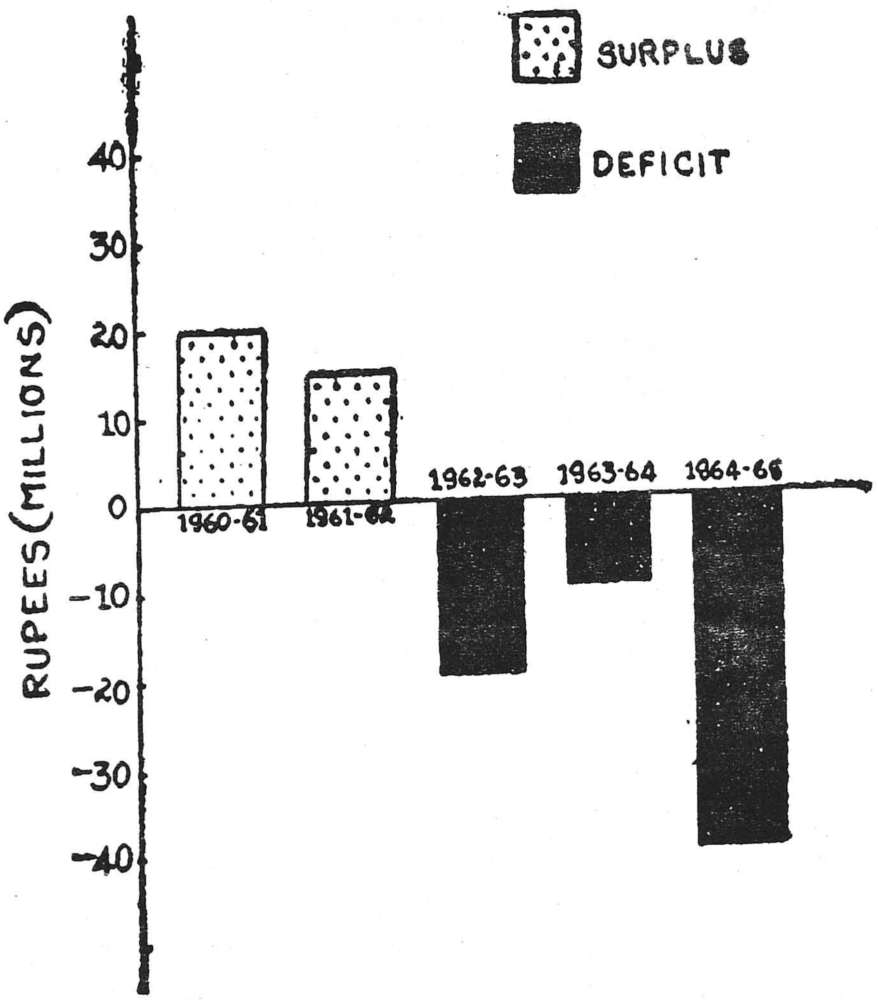
| Year | Exports | Imports | Balance of Trade (Deviation) | |
|---|---|---|---|---|
| | | | + | - |
| 1960-61 | 120 | 100 | 20 | |
| 1961-62 | 100 | 85 | 15 | |
| 1962-63 | 70 | 90 | | 20 |
| 1963-64 | 120 | 130 | | 10 |
| 1964-65 | 140 | 180 | | 40 |

Fig : 10

Value of indo - U.S. Trade During 1960 - 65

Scale : 1 cm = Rs. 10 millions

Index




## RECTANGULAR AND SQUARE DIAGRAMS

### Area diagrams

In the case of bar diagrams, we have used length to represent the given magnitudes. As against this, area is also to be used to represent the given magnitudes. Such diagrams are known as Area diagrams. Important types of area diagrams are as follows :

(i) Rectangular diagrams

(ii) Square diagrams

(iii) Pie diagrams

### Rectangular diagrams

A rectangle has two sides viz., length and breadth and are taken into account while representing data. When two sets of figures are given they are represented by rectangle as follows.

The ratio of the totals of the two sets of figures is found out first and the widths of the two rectangles are taken in that ratio.

Total of the figures in each set is equated to 100 and each figure in each set is calculated as percentage of the total. Now the lengths of each of the two rectangles is taken as 100 percent and it is subdivided into various components.

Let us consider the following table.

**Illustration - 11**

| | | First Plan Rs. (crores) | Second Plan Rs. (crores) |
|---|---|---|---|
| 1) | Agriculture and Community Development | 357 | 568 |
| 2) | Irrigation and power | 661 | 913 |
| 3) | Industry and mining | 179 | 890 |
| 4) | Transport and communication | 5571 | 35 |
| 5) | Social Services | 533 | 945 |
| 6) | Miscellaneous | 69 | 99 |
| | Total | 2356 | 4800 |

We represent the above Table by means of rectangular diagram as follows :

Total expenditure in the First Plan is Rs. 2356 crores and in the Second Plan is Rs. 4800 crores. The ratio of these two totals is 2356 : 4800. This 1 : 2 (approximately). Now two rectangels representing the expenditures in the First and Second Plans respectively are drawn such that their widths are in the ratio 1 : 2. That is, if the first rectangle is drawn with 1 cm width then the second rectangle is drawn with 2 cms as width.

The heights of the two rectangles are taken to be the same. That is the first rectangle is drawn with 10 cm as height then the second rectangle is also drawn with 10 cm as height.

For each Plan, the expenditure under each head is calculated as a percentage of the total expenditure. Now just as in the case of percentage bar diagrams, each rectangle is divided into various subdividions, each subdivision representing the percentage expenditures under each head. In the table below we have given the calculated percentage expenditure under each head.

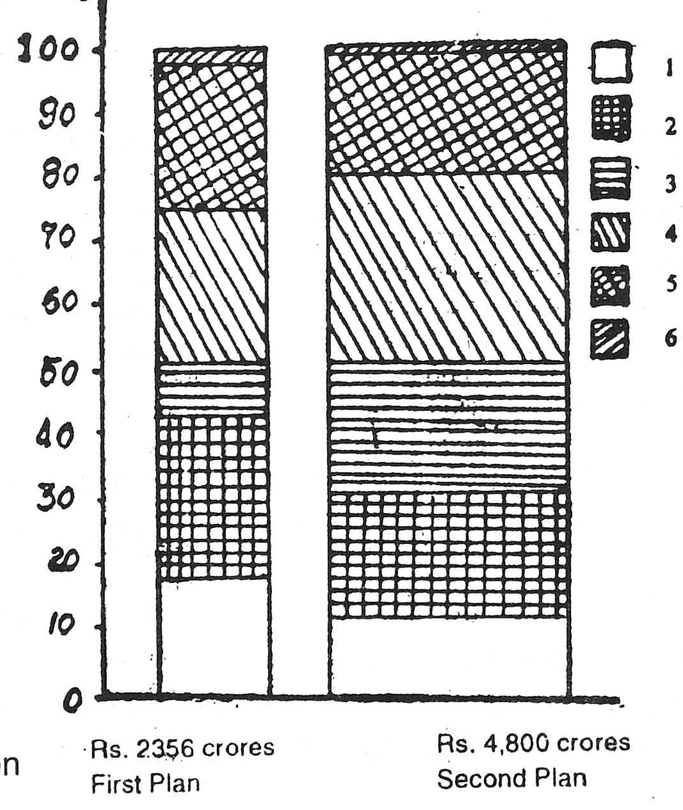
## Plan Expenditure

| Heads of Expenditure | First Plan Rs. (crores) | First Plan % Rs.2356 crs = 100 | Second Plan Rs. (crores) | Second Plan % Rs.480 crs = 100 |
|---|---|---|---|---|
| Agriculture and Community devlelopment | 357 | 15.1 | 568 | 11.8 |
| Irrigation and Power | 661 | 28 | 1913 | 19.0 |
| Industry and Mining | 179 | 7.6 | 890 | 18.5 |
| Transport and Communication | 557 | 23.6 | 1385 | 28.9 |
| Social Services | 533 | 22.6 | 945 | 19.7 |
| Miscellaneous | 69 | 3.0 | 99 | 2.1 |
| Total | 2356 | 100 | 4800 | 100 |

The Rectangular Diagram is as follows :

**Fig : 11**

**Distribution of outlay in the first two Five Year Plans**




1. Agriculture and Community
2. Irrigation and Power
3. Industry and Mining
4. Transport and Communication
5. Social Services
6. Miscellaneous

These types of comparisons are possible with the help of a rectangular diagram. They are as follows :

1. In the above diagram, the areas of the two rectangles are in proportion to the total expenditures given. Hence, the change in the total expenditure can be easily found out by comparing the areas of the two rectangles drawn. Thus, comparison of the given totals is possible with the help of rectangular diagram.

2. The areas of the components of the rectangles are in proportion to the actual expenses (given in Rs.) under various heads. Hence a comparison of the areas of the components in the first rectangle will give us the change in the actual expenditure made under different heads in the First Five Year Plan. Similarly, the change in the actual expenditure made under different heads in the Second Five Year Plan can be found out by comparing the areas of the components in the second rectangle. In addition to these the actual expenditure made under each head in the two plans can also be compared by comparing the areas representing the same component in the two rectangles. In short, the area of the components of the rectangles facilitate comparison of actual expenditure made under different heads in the same Plan and in the two plans.

Thus, rectangular diagram facilitates comparison of absolute values of the components.

3. Percentage value of each sub-division is represented by heights of each component of the rectangles. Hence, we can make a comparison of the percentage values of subdivisions of the given total by comparing the heights of the components of rectangles.

In our diagram, the height of the first component in the first rectangle is greater than that in the second rectangle. This implies that the expenditure made on agriculture and community development forms a greater percentage of total expenditure in the I Plan compared to the II Plan. Similarly, we can make comparisons of the remaining components also.

In the above example, we have used the rectangles to show actual as well as percentage distribution of outlay in the First and Second Five Year Plans. Here the heights of the rectangles are equal; only the widths are different. But by varying the height of the rectangle also, the rectangular diagram can be used to represent three different factors. For example, suppose we are given data relating to sale, proceeds, quantity produced and different expenses of production of a commodity for two years of two different commodities for given year. Here rectangular diagram can be used to show (1) price of the commodity or commodities (2) quantity produced and (3) different expenses of production and net profit.

Consider the following example :

**Illustration - 12**

| | Commodity A | Commodity B |
|---|---|---|
| Price per units Rs. | 10 | 12 |
| Quantity sold Rs. | 20 | 24 |
| Value of raw materials Rs. | 100 | 120 |
| Other expenses Rs. | | |
| Profits Rs. | | |

We represent the above data by means of rectangular diagram as follows :

We are given prices, quantities sold and different expenses of production and profits, for two commodities A and B. We draw a rectangle for each commodity. We take the heights of the rectangles in the ratio of the price of two commodities, namely 10 : 12. That is, if we draw a rectangle with 10 cms as its height to represent the commodity A, then a rectangle with 12 cms as height is drawn to represent B. The width of the two rectangles are taken in the ratio of quantity sold of A and B (i.e) in the ratio 20:21 (ie) in the ratio 5:6. That is, if we take 5 cms as the width of the first rectangle, then we take 6 cms as the width of the second rectangle.

After drawing the two rectangles, we have to mark the different expenses and profit inside these rectangles as components. In the table we are given total expenses and total profits. But total cost per unit added with profits per unit gives price per unit of each commodity. Hence, we calculate the cost per unit and profit per unit for A and B as follows :

$$\text{Raw material cost per unit for A} = \frac{\text{Total cost of raw materials for A}}{\text{Quantity sold of A}}$$

$$= \text{Rs. } \frac{100}{20}$$

$$\text{Other costs per unit for A} = \frac{\text{Other costs in total for A}}{\text{Quantity sold for A}}$$

$$= \text{Rs. } \frac{60}{20} = 3$$

$$\text{Profit per unit for A} = \frac{\text{Total profits for A}}{\text{Quantity sold of A}}$$

$$= \text{Rs. } \frac{40}{20}$$

$$= \text{Rs.}2$$

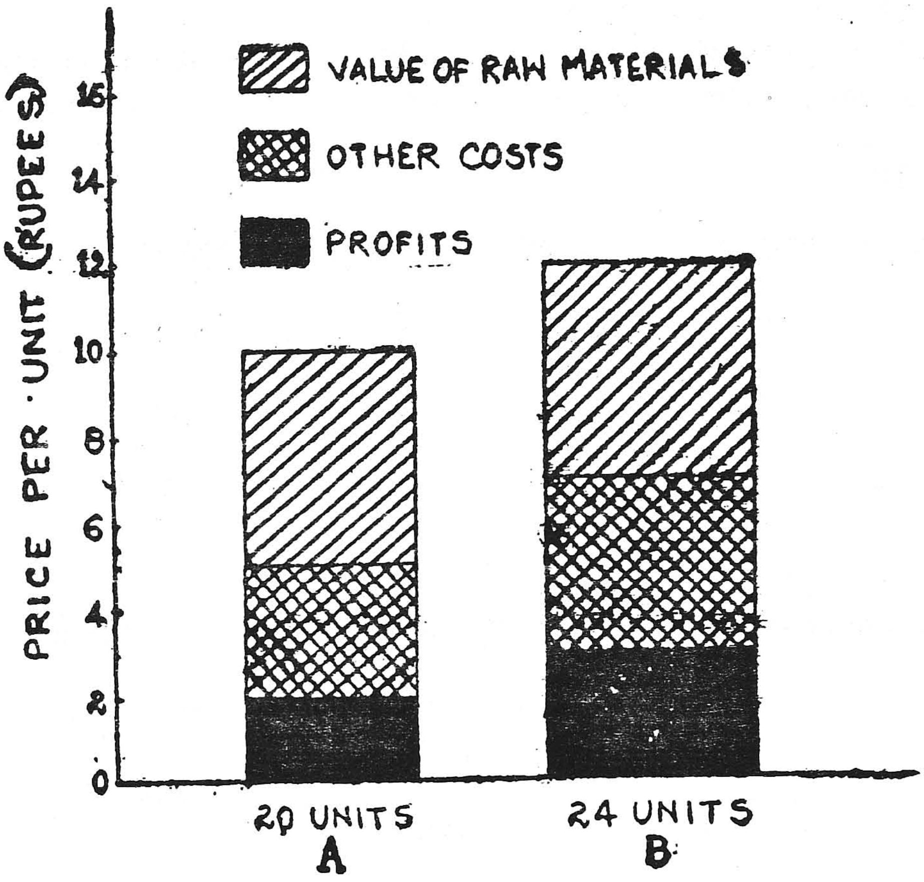
In the same way, we can calculate for B also. We have the unit costs profits for A and B in the following table:

| | A | B |
|---|---|---|
| Value of Raw materials per unit Rs. | 5 | 5 |
| Other expenses per unit Rs. | 3 | 4 |
| Profit per unit Rs. | 2 | 3 |

In marking the subdivisions we follow the same procedure as in the case of Bilateral bar diagram. Now, the rectangular diagram is as follows :

**Fig : 12**

**DETAILS OF COST OF TWO COMMODITIES A AND B**




In the above diagram, total area of each rectangle represents total sale proceeds from each commodity. Areas of the components of each rectangle represent the total value of raw materials, total value of other costs and total profits for each commodity. The total height of each rectangle shows the price per unit of each commodity and the heights of the components show the value of raw materials per unit, other costs per unit and profit per unit.

## Special merit of rectangular diagram

Rectangles are the only two dimensional diagrams which are capable of showing three different factors. Though squares and circles are also two dimensional diagrams they do not enjoy this property. Hence, the place of rectangles in a diagrammatic representation is very important.

## Pie diagrams

When a set of data are given, they can be represented by circle whose areas are proportional to the given quantities. This representation is known as pie diagram or circular diagram.

A circle can be drawn only if we know the radius of the circle. So, if the given magnitudes are to be represented by means of areas of circles, first we must decide about the radius to be taken to draw the circles.

The area of a circle varies as the square of its radius. Hence, if the areas of a circle is to be proportional to a given magnitude, the radius of the circle must be proportional to the square root of the given magnitude. For example, if 4 and 25 are two given magnitudes to be represented by areas of two circles, we must draw the circles in such a way that their radius are in the ratio $\sqrt{4} : \sqrt{25} = 2:5$. That is if we denote, 4 by a circle with 2cms as its radius, then 25 is represented by a circle with 5 cms as its radius.

The circles are generally drawn such that their centres lie on a straight line. Also equal space is left in between any two circles.

## Illustration - 13

Number of students who appeared for B.A., examination at centres A, B and C are as follows :

| Centre | Total number appeared |
|---|---|
| A | 400 |
| B | 625 |
| C | 900 |

The above data can be represented by the areas of three circles whose radius are in the ratio

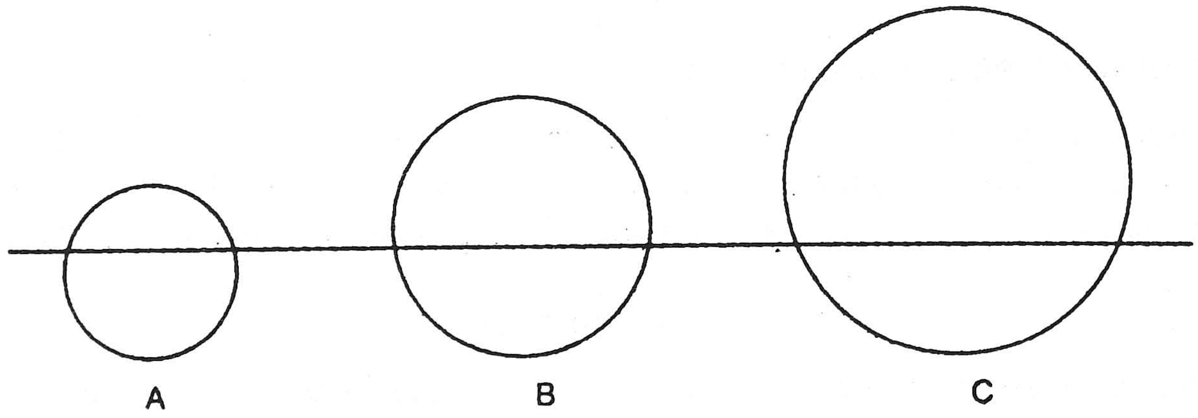
$$\sqrt{400} : \sqrt{625} : \sqrt{900}$$

$$= 20 : 25 : 30$$

$$= 4 : 5 : 6$$

**Fig : 13**

**NUMBER OF STUDENTS WHO APPEARED FOR B.A., EXAMINATION**




**AT THREE CENTRES A, B AND C**

The above data can as well be represented by squares also taking the sides of the squares in the ratio 4 : 5 : 6. But circles are easy to draw compared to squares. Also, circles look more

beautiful and attractive to look at and in many cases more effective to compare than squares. Therefore, when there is a choice between squares and circles, the latter are generally preferred.

## Angular diagram

If the given data consist of a total and its subdivision they can also be represented by means of a pie diagram. Here the given total is represented by the area of a circle of any radius and the subdivisions are represented by areas of sectors cut inside this circle. That is, we draw a circle with some radius to represent the given total ; inside the circle various sectors are cut in such a way that their areas are in the ratio of magnitudes of the various subdivisions of the given total. The area of a sector is proportional to its angle's and also, the sector can be drawn only if we know is angle.

So in order to represent the given subdivision by means of area of sectors, first we have to convert the values of sub-division into angles. For this conversion, we adopt the following procedure.

The given total figure is represented by the total area of the circle. In other words, in terms of angle, the given total is represented by the total angle around the centre. The total angle around the centre is equal to 360°. Therefore, by equating the given total to 360° we calculate the value of subdivision in terms of degrees. For example, consider the following table.

## Illustration - 14

The following table gives the monthly expenditure of family A.

| Items of expenditure | Expenditure in Rs. |
|---|---|
| Food | 105 |
| Clothing | 65 |
| House Rent | 50 |
| Education | 35 |
| Fuel and lighting | 20 |
| Miscellaneous | 25 |
| Total | 300 |

1. If we draw two radii inside a circle the area included between the two radii is called a 'sector'.
2. The angle included between the two radii of a sector is called its angle.

By definition the total angle around a point is always equal to 360° . Centre of a circle is also a point. Hence, the total angle around the centre is equal to 360°.

To represent the above table by means of Pie diagram first we equate the total expenditure of Rs.300 to $360^0$ and then we calculate the value of each subdivision in terms of degrees

Expenditure on food in terms of degrees $= \dfrac{\text{Expenditure on food in Rs.}}{\text{Total expenditure in Rs.}} \times 360$

$= \dfrac{105}{300} \times 360$

$= 126$

Expenditure on clothing in terms of degrees $= \dfrac{\text{Expenditure on clothing in Rs.}}{\text{Total expenditure in Rs.}} \times 360$

$= \dfrac{65}{300} \times 360$

$= 78$

Similarly, we can calculate the values of the remaining items of expenditure also in terms of degrees.

Value of each item of expenditure expressed in degrees gives us the angle of each sector to be cut inside the circle. In the following table we have given the values of all the items in degree under the heading 'Angle of Sector'.

| Items of expenditure | Expenditure (in Rs.) | Angle of Sector (in degrees) |
|---|---|---|
| Food | 105 | $\dfrac{105}{300} \times 360 = 126$ |
| Food | 105 | $\dfrac{105}{300} \times 360 = 126$ |
| Food | 105 | $\dfrac{105}{300} \times 360 = 126$ |
| Food | 105 | $\dfrac{105}{300} \times 360 = 126$ |
| Food | 105 | $\dfrac{105}{300} \times 360 = 126$ |
| Food | 105 | $\dfrac{105}{300} \times 360 = 126$ |
| Total | 300 | 360 |

Knowing the angles, the sectors can be easily drawn as follows : First we draw some radius, usually, the vertical radius above the centre. Next we place the protractor* in such a way that its base line coincides with the vertical radius drawn and the edge of the arrow mark at the base of the protractor coincides with the centre of the circle. Now, we measure on angle of $126^0$ and draw a radius. The area of the sector formed by this radius and vertical radius represents the expenditure on food by family A.

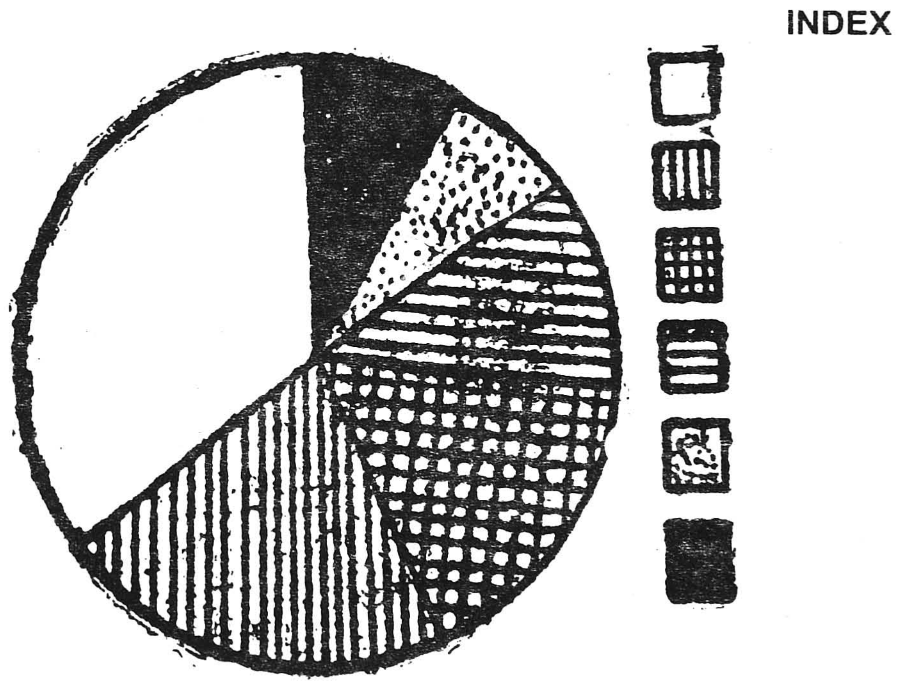
To mark the second item namely, clothing, we draw another sector adjacent to the first sector drawn. The angle of this second sector must be equal to $78^0$. To draw this sector protractor is placed in such a way that its base line coincides with the radius with which the first sector ends and edge of the arrow mark at the base of the protractor coincides with the centre of the circle. After placing the protractor in this way, an angle of $78^0$ is measured from the base line and a radius is drawn. The area of the sector formed by this radius and the radius with which, the first sector ends represents the expenditure on clothing.

We follow the same procedure to mark the remaining items also by means of sectors.

Usually, the items are marked inside the circle, in the anti-clockwise direction. Also, the items are given in descending order of magnitude. However, the miscellaneous item is given last even if it is larger than some of the other items. When there is sufficient space, wording can be given horizontally inside the sectors. The lettering should never be radiating from the centre nor placed along the circumference. Distinct colours or shadings may also be used to emphasise the divisions.

**Fig : 14**

**MONTHLY EXPENDITURE OF FAMILY - A**




---

* The instrument which is used to measure angles is called protractor.

## Angular diagram for two sets of data

If we are given two totals and their subdivisions they can also be shown by angular diagram. Here, two totals are represented by two circles whose areas are in the ratio of the given totals. For this, we must take their ratio in the ratio of the square roots of the two totals given. For example, if 100 and 400 are the given totals, than they are represented by the areas of two circles whose radii are in the ratio 100 : $\sqrt{400}$ ≠ 10 : 20 = 1:2

After drawing the circles, the subdivisions are marked inside each circle as usual. The items are shown in the same order in the two circles.

### Illustration - 15

Details of monthly expenditure of two families. A and B are in the following table.

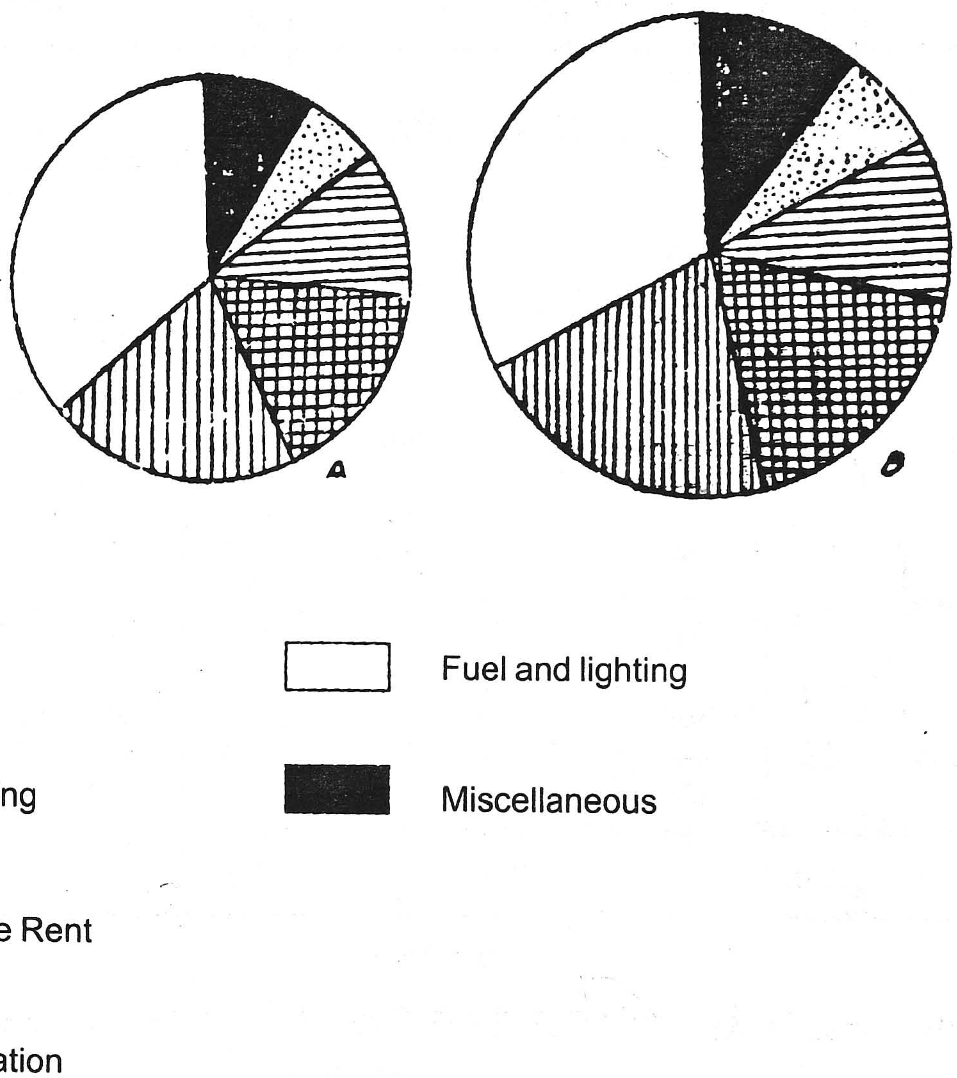
| Items of expenditure | Family A | | Family B | |
|---|---|---|---|---|
| | Expenditure in Rs. | Angle of Sector (in degrees) | Expenditure in Rs. | Angles of Sector (in degrees) |
| 1) Food | 105 | $\frac{360}{300} \times 105 = 126$ | 115 | $\frac{360}{300} \times 115 = 115$ |
| 2) Clothing | 65 | $\frac{360}{300} \times 65 = 78$ | 75 | $\frac{360}{300} \times 76 = 75$ |
| 3) House Rent | 50 | $\frac{360}{300} \times 50 = 60$ | 65 | $\frac{360}{300} \times 65 = 65$ |
| 4) Education | 35 | $\frac{360}{300} \times 35 = 42$ | 40 | $\frac{360}{300} \times 40 = 40$ |
| 5) Fuel and lighting | 20 | $\frac{360}{300} \times 20 = 24$ | 25 | $\frac{360}{300} \times 5 = 25$ |
| 6) Miscellaneous | 25 | $\frac{360}{300} \times 25 = 30$ | 40 | $\frac{360}{300} \times 40 = 40$ |
| Total | 300 | 360 | 360 | 360 |

Ratio of the square root of the total expenditure is $\sqrt{300}$ : $\sqrt{360}$ i.e., 5.5:6 (approximately).

∴ The two circles are drawn such that their radii are in the ratio 5.5 : 6.

**Fig : 15**

**MONTHLY EXPENDITURE OF THE FAMILIES A AND B**




Percentage Pie diagram

Sometimes we may be given two sets of data, each set consisting of components expressed as percentages of the whole. Here, as each of the given total is equal to 100%, we draw circles of equal radii to represent the given totals. Then, in the usual manner, the components are shown by sectors cut inside each circle.

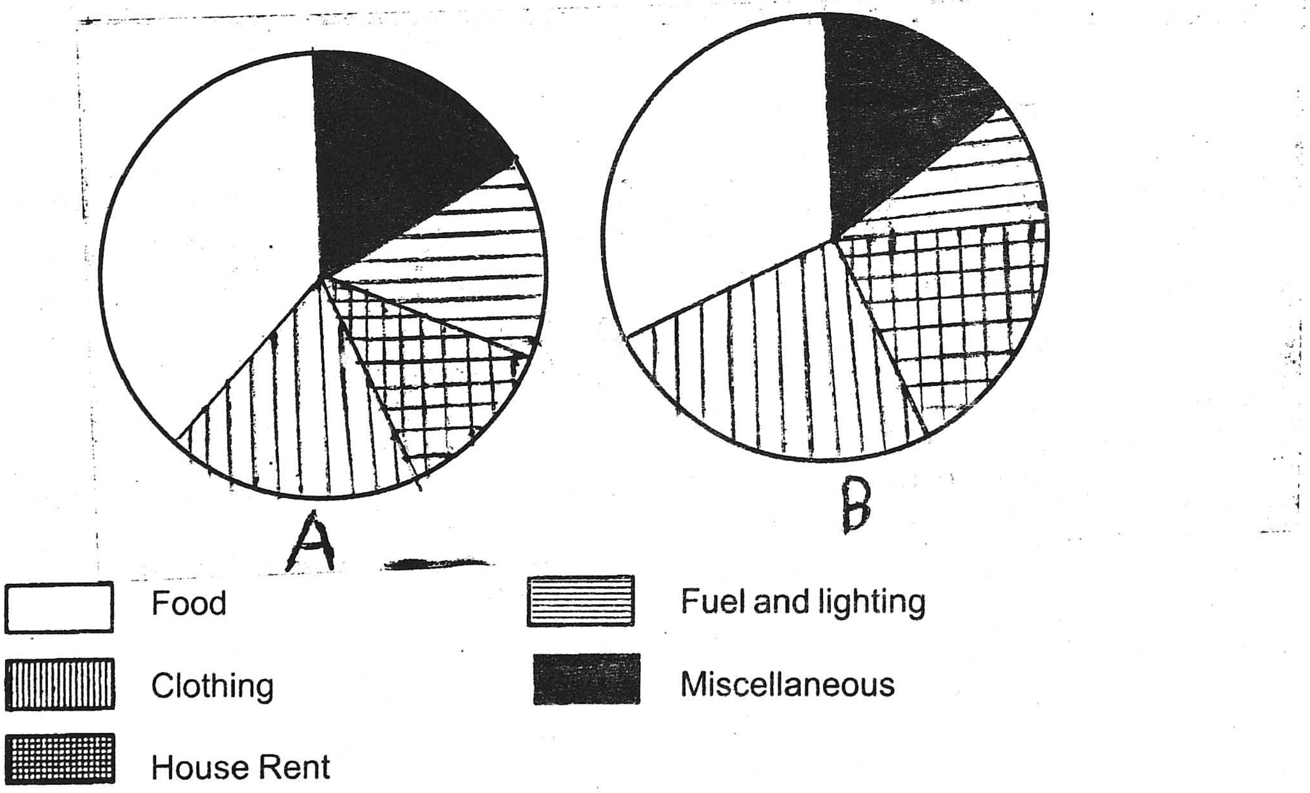
## Illustration - 16

The following Table gives the monthly expenditures of two families A and B under different heads as percentages of the whole. Draw Percentage Pie diagram.

| Items of expenditure | Family A | | Family B | |
|---|---|---|---|---|
| | Expenditure in Rs. | Angle of Sector (in degrees) | Expenditure in Rs. | Angles of Sector (in degrees) |
| 1) Food | 5 | $\frac{360}{300} \times 35 = 126$ | 30 | $\frac{360}{300} \times 30 = 108$ |
| 2) Clothing | 20 | $\frac{360}{300} \times 20 = 72$ | 25 | $\frac{360}{300} \times 25 = 90$ |
| 3) House Rent | 15 | $\frac{360}{300} \times 15 = 54$ | 20 | $\frac{360}{300} \times 20 = 72$ |
| 4) Fuel and lighting | 0 | $\frac{360}{300} \times 10 = 36$ | 10 | $\frac{360}{300} \times 10 = 36$ |
| 5) Miscellaneous | 20 | $\frac{360}{300} \times 20 = 72$ | 15 | $\frac{360}{300} \times 15 = 54$ |
| Total | 100 | 360 | 100 | 360 |

**Fig : 16**

**MONTHLY EXPENDITURE OF TWO FAMILIES A AND B**



## Usefulness of pie diagram

Pie diagrams are popularly used in representing statistical facts. The reason for their popularity lies in the facility and ease with which they can be drawn. The propagandist makes the best use of Pie diagrams. Pie diagram appeals to the popular mind especially in representing financial data. As an example, let the expenses of a labourer on various items like food, clothing, drink, recreation be represented by a Pie diagram. It will be appealing to the labourer the larger portion of his income is wasted on drink.

Pie diagram is a good diagram for showing the distribution of world population into various continents or of world area into areas for different continents.

## Limitations of Pie diagrams

Although Pie diagrams are widely used, sub-divided bars or subdivided rectangles are preferred because of the following reasons :

1) It is easier to draw subdivided bars or subdivided rectangles than to draw subdivided circles. Because, the calculation of angles of sectors involves considerable labour and time.

2) The heights of subdivisions of rectangles or bars are measurable in the scale showing percentages. But the sectors of angle directly in percentages. But the sectors of circles are comparable directly in terms of angles and only indirectly in term of percentages.

3) An optical illusion may result in going round the sectors.

## Pictograms

In pictograms relative values of items are represented by actual pictures. The number of pictures drawn are in proportion to the various items which have to be represented. Let 3,500 7,000 and 8,500, represent the number of cars produced by a country in the three years. 1950 to 1952. We can represent this by 7,14 and 17 small equal size cars respectively drawn in neat row each car representing the production of 500 cars. Sometimes the above data can be represented by a single car for each year, the size of the car being proportional to the magnitude. But it is far better to represent magnitudes not by sizes of pictures, but by rows of similar pictures of the same size. The reason is that it is far easy to compare lengths better than areas or volumes when they are pictorically represented.

The pictogram is a very popular form of appeal to the public and is almost sure to attract attention. It is therefore useful for exhibition work. Official statistics often contain pictograms.

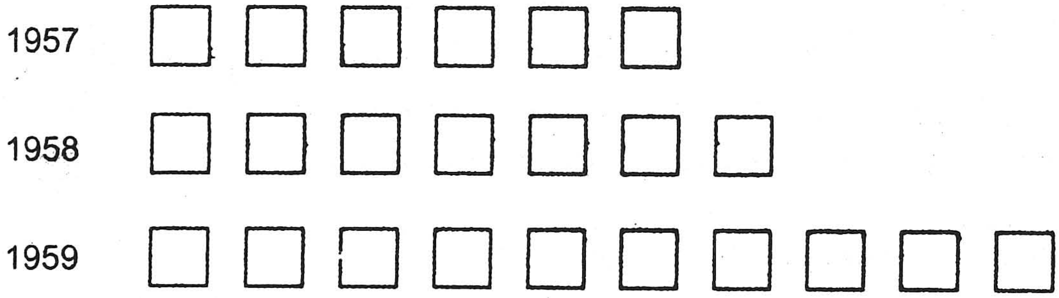
This method of representation is frequently used in backward areas where level of literacy is low and people are unable to interpret figures. But the disadvantage of this method is that generally people cannot draw good pictures.

### Illustration - 17

The following Table gives the production of tea in three years 1957 to 1959

| Year | Production of tea (in lakhs of tons) |
|---|---|
| 1957 | 6 |
| 1958 | 7 |
| 1959 | 10 |

We can represent this by drawing 6 tins in the 1st row; 7 tins in the second row and 10 tins in the 3rd row all the tins being of the same size. Each tin represents one lakhs tons of tea.




## Choice of a suitable diagram

The choice of a diagram depends upon the nature of the data and the type of people for whom the diagram is meant. Whether to choose bar diagrams and if bar diagrams are chosen whether to draw simple bars or component bars or percentage bars all depend upon the nature of data given. The type of people for whom the diagram is intended must also be taken into account. For instance to attract the attention of an uneducated mass, pictograms are more suitable than rectangles, squares, circles, etc. Different types of diagrams have specialised uses of follows :

a) Whenever changes in totals alone are to be represented, simple bar diagram should be used. For example, to show the differences in the production of rice in various states of India, to show the total population in the last five decades, etc., simple bar diagram should be used.

b) Whenever changes in the aggregate as well as the size of the components are to be exhibited, component or subdivided bar diagram should be used. For example to show the changes in the total expenditure as well as in expenditure under each head in two five year plans, component bar diagram should be used.

c) Whenever changes in the values of component parts are to be shown and where changes in the aggregate are of no importance multiple bar diagram is suitable.

d) Where changes in the relative values of components alone are to be shown, percentage bar diagram is suitable.

However, multiple and component bar diagrams are used only when there are not more than three or four components. When a large number of components have to be shown, a pie chart is more suitable. The effectiveness of pie diagram also tends to dwindle with more than seven or eight components.

When more than two or three sets of data are to be compared, pie diagram should not be used as the areas of a number of circles cannot be compared easily.

Rarely, areas of circles are used to represent size. If it is possible to use bars, then circles should be avoided as the eye can easily compare lengths than areas.

When changes in aggregators and changes in absolute as well as relative values of components are to be displayed, rectangles are the only types of diagram to be used.

Pictograms are more informative and more effective forms of representation of data to the general public who, by and large neither possess much ability to understand nor take interest in the less attractive forms of presentation.

## Model Questions

1) Explain the usefulness of diagrams, charts and graphs in a research report.

2) Write a note on drawing bar and pie diagrams.

## LESSON - 9

# METHODS OF PRESENTATION OF STATISTICAL INFORMATION - II
# FREQUENCY GRAPHS

Graphic representation of a frequency distribution is called frequency graph. There are various types of frequency graph. The most important types are as follows:

(i) Histograms

(ii) Frequency Polygon

(iii) Frequency curve and

(iv) Ogives

## HISTOGRAMS

Histogram consists of a continuous series of rectangles drawn over the true class intervals of a frequency distribution. The area of each rectangle is used to represent the frequency of the respective class interval. Therefore, total area of all the rectangles, taken together will represent the total frequency. As the rectangles are found without any gap, histogram is some times called 'column diagram'.

Now, let us see the step by step procedure of drawing a histogram (i) in the case of a distribution having equal class intervals and (ii) in the case of a distribution having unequal class intervals.

Histogram for equal class Intervals :

(i) First, we select a suitable scale and mark the true class limits of the given frequency distribution on the X-axis.

(ii) Taking another convenient scale the frequencies are marked on the Y-axis. It is not essential to show the zero-point on the X-axis. But the scale on Y-axis must start with zero.

(iii) Now, the bars are to be erected. Having the distance between the true class limits of each class interval as base and the respective class frequency as height, various rectangles are drawn. As many rectangles are drawn as there are class intervals in the given frequency distribution . Since the given distribution has equal class intervals, widths of all the rectangles will be equal. Therefore, in this case where the distribution is having equal class intervals, the area as well as the height of each rectangle is proportional to the respective class frequency. The total area of all the rectangles will be proportional to the total frequency.

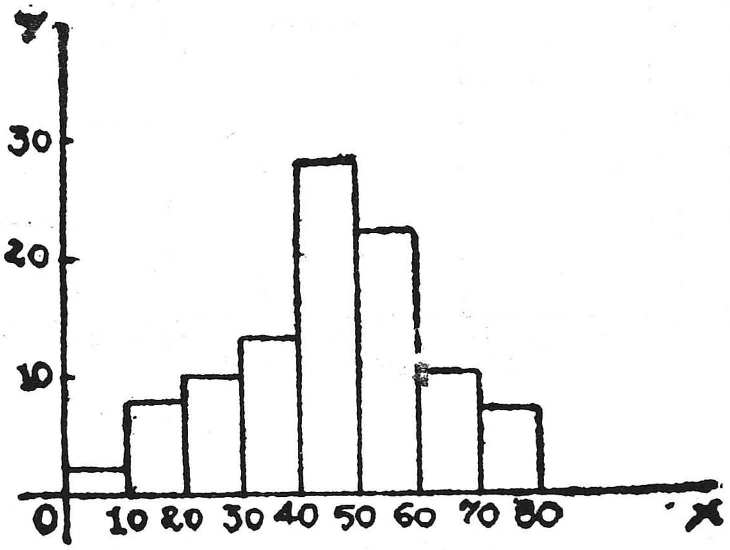
Consider the following examples.

### Illustration -1

Distribution of marks of 100 students in an examination at Madurai centre is given below :

| Marks | Numbers of Students |
|---|---|
| 0-10 | 2 |
| 10-20 | 8 |
| 20-30 | 10 |
| 30-40 | 13 |
| 40-50 | 28 |
| 50-60 | 22 |
| 60-70 | 10 |
| 70-80 | 7 |
| Total | 100 |

In this problem, we are already given true class intervals. So, we mark the given class intervals as they are on the X-axis, taking the scale 1 cm = 10 marks. We mark the frequency on the Y -axis taking the scale 1 cm = 5 students. Now the histogram is as follows.




### Illustration -2

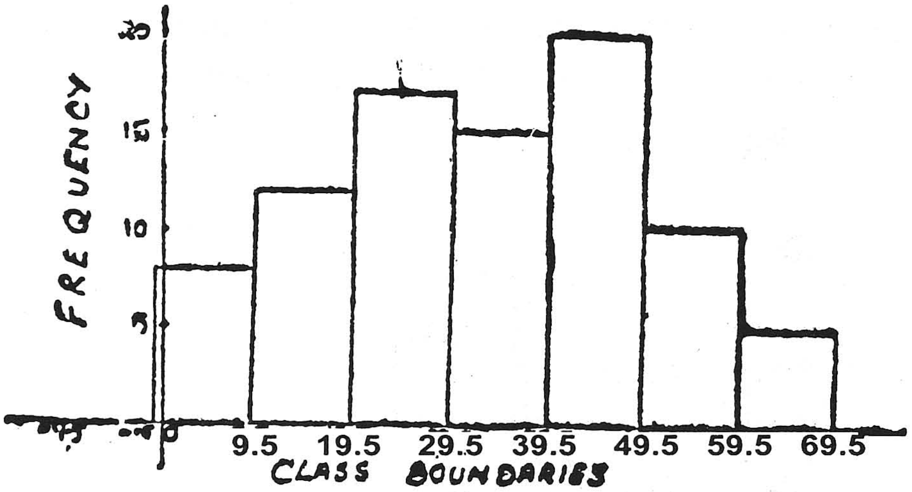
Draw a histogram from the following data.

| Class intervals | Frequency |
|---|---|
| 0-9 | 8 |
| 10-19 | 12 |
| 20-29 | 17 |
| 30-39 | 15 |
| 40-49 | 20 |
| 50-59 | 10 |
| 60-69 | 5 |

Here, the given class intervals are not true class intervals. So first we convert them into true class intervals and give them as follows :

| True class intervals | Frequency |
|---|---|
| .5 - 9.5 | 8 |
| 9.5 - 19.5 | 12 |
| 19.5 - 29.5 | 17 |
| 29.5 - 39.5 | 15 |
| 39.5 - 49.5 | 20 |
| 49.5 - 59.5 | 10 |
| 59.5 - 69.5 | 5 |

We have given below the histogram




Sometimes, we may be given mid-values instead of class intervals of the frequency distribution. In such cases also, first of all we must find out the true upper and true lower limits of each class and draw the histogram as before.

## Histogram for unequal class intervals

Ordinarily , frequency distributions are constructed with class intervals of equal magnitude. But to meet special situations a frequency distribution with unequal class intervals may have to be constructed. To represent such a distribution by means of histogram, it is necessary to make adjustments for the varying lengths of class intervals. These adjustments are made by finding out frequency densities. Now let us see the meaning and the method of calculation of frequency densities.

Suppose, in the given frequency distribution, 10 is the length of most of the class intervals. Now, this length 10 is taken as the standard length and the frequencies of the remaining classes

(whose lengths are either greater than or less than 10 are converted in terms of the class interval of 10. That is, frequency of each class is divided by the corresponding length and multiplied by 10. The resultant figures are called, 'frequency densities'.

Now, true class intervals are marked on the X-axis and the frequency densities are marked on the Y -axis.

Taking the true class intervals as the widths and the corresponding frequency densities as the heights, series of rectangles are drawn. This continuous series of rectangles give the histogram of the given distribution.

Here, only the area of each rectangle and not its height is proportional to the respective class frequency. Total area of all the rectangles is proportional to the total frequency.

Consider the following example.

## Illustration - 3

Average weekly earnings of the employees in a factory are as follows :

| Weekly earnings | Frequency |
|---|---|
| 50-55 | 5 |
| 55-60 | 9 |
| 60-65 | 17 |
| 65-70 | 23 |
| 70-75 | 47 |
| 75-80 | 58 |
| 80-90 | 62 |
| 90-100 | 76 |
| 100-110 | 96 |
| 110-120 | 36 |
| 120-140 | 32 |

In the above Table, the first six classes are having the same width of 5 next four classes of 10 and the last class 20.

Here, majority of the classes have the common length namely 5. Hence, 5 is taken as the standard length. The frequencies of the remaining classes are converted in terms of the class interval of 5 and the frequency densities are obtained as in the following Table :

| Weekly earnings in Rs. | Frequency | Frequency density, number of employees per Rs. 5 of earnings |
|---|---|---|
| 50-55 | 5 | 5 |
| 55-60 | 9 | 9 |
| 60-65 | 17 | 17 |
| 65-70 | 23 | 23 |
| 70-75 | 47 | 47 |
| 75-80 | 58 | 58 |
| 80-90 | 62 | $\frac{62}{10} \times 5 = 31$ |
| 90-100 | 76 | $\frac{76}{10} \times 5 = 38$ |
| 100-110 | 96 | $\frac{96}{10} \times 5 = 48$ |
| 110-120 | 36 | $\frac{36}{10} \times 5 = 18$ |
| 120-140 | 32 | $\frac{32}{20} \times 5 = 8$ |

Now, as usual, the true class intervals are marked on the X-axis. The frequency densities are marked on the Y-axis. Having the true class interval as the base and the corresponding frequency density as the height rectangles are drawn. The resultant figure is the required Histogram and it is as follows :

**WEEKLY EARNINGS OF EMPLOYEES IN A FACTORY**

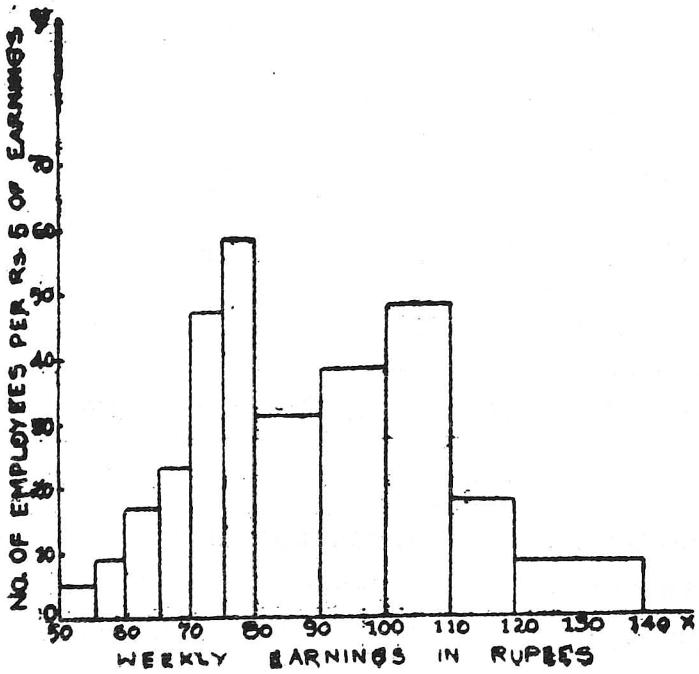


**Illustration - 4**

The following table gives the average daily sale of 500 firms. Draw the histogram.

| Sales in Rs. | Number of firms |
|---|---|
| 0 - 500 | 10 |
| 500 - 1,000 | 25 |
| 1,000 - 2,000 | 53 |
| 2,000 - 3,000 | 100 |
| 3,000 - 4,000 | 140 |
| 4,000 - 5,000 | 152 |
| 5,000 above | 20 |

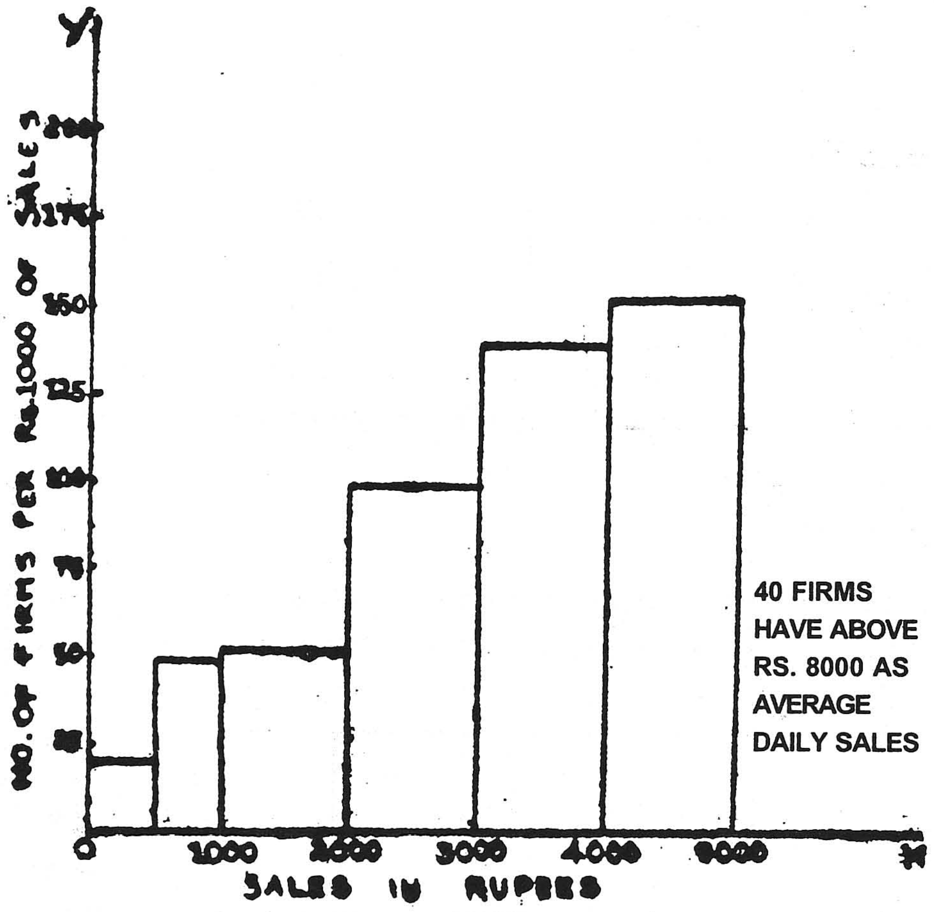
In the above Table, the firs ttwo classes are having the same width, namely 500. The next four classes are having the same width, namely 1,000. The last class is an open class and hence its length is indeterminate.

The most common width in the given frequency table is 1,000 and hence, it is taken as the standard width and the frequency densities are calculated. As the width of the last class is indeterminate, the frequency density for it cannot be determined.

| Sales in Rs. | Number of firms | Frequency density of firms (per Rs. 1,000 of sales) |
|---|---|---|
| 0 - 500 | 10 | $\frac{10}{500} \times 1000 = 20$ |
| 500 - 1,000 | 25 | $\frac{25}{500} \times 1000 = 50$ |
| 1,000 - 2,000 | 53 | 53 |
| 2,000 - 3,000 | 100 | 100 |
| 3,000 - 4,000 | 140 | 140 |
| 4,000 - 5,000 | 152 | 152 |
| 5,000 above | 20 | - -- |

Now the true class intervals are marked on the X-axis and the frequency densities on the Y -axis. As usual, with class intervals as widths and the corresponding frequency densities as heights rectangles are drawn and Histogram is obtained.

## AVERAGE DAILY SALES OF FIRMS




(Please note how the frequency of the last class interval (which is an open class) is represented in the above Histogram)

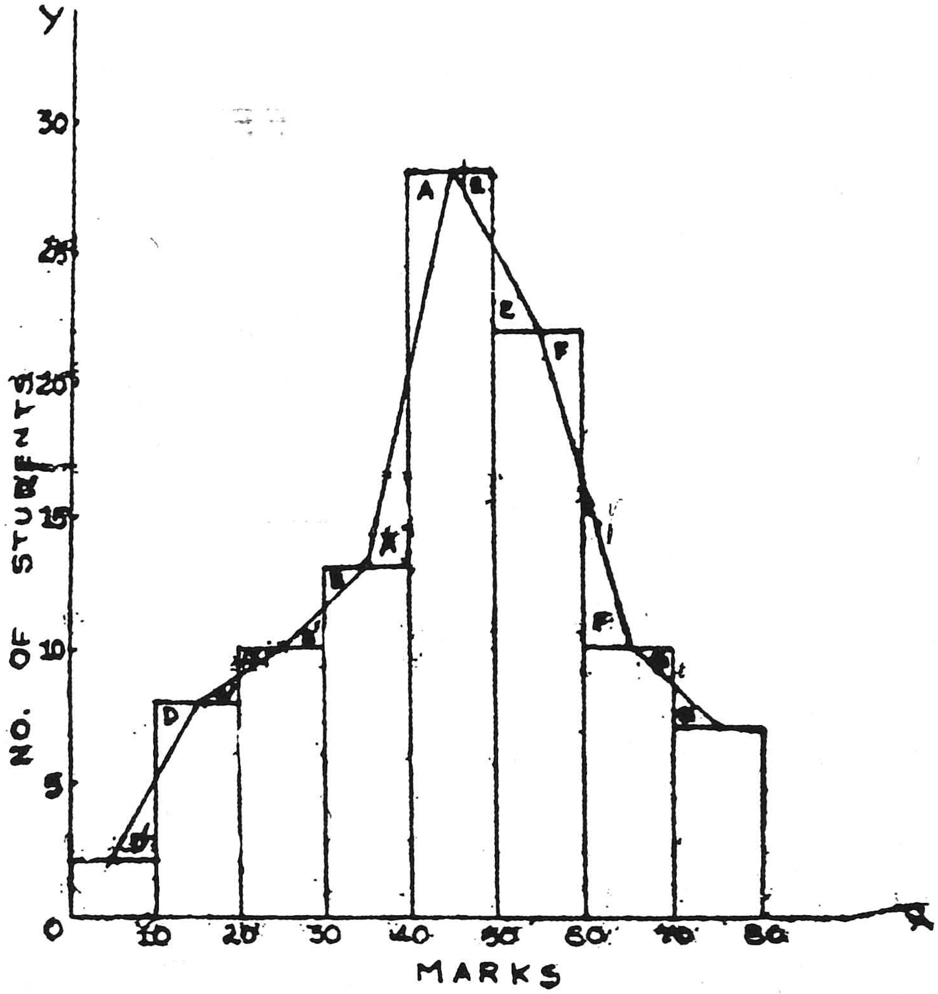
## FREQUENCY POLYGON

If we join the midpoint of the top of each rectangle with that of the adjacent rectangle of a histogram, the resultant graph is known as Frequency Polygon. This graph is drawn on the assumption that the values in a class are evenly distributed throughout the class. The mid-value of a class is consequently taken as representing the class.

Consider the following example

## Illustration - 5

Consider the same data given under illustration I and draw the frequency polygon.




As explained in illustration 1, first we draw the histogram. In this histogram we marked the midpoint of the top of each rectangle. These midpoints are joined by means of straight lines as shown in the following diagram and the Frequency Polygon is obtained.

It is to be noted in the above diagram that the frequency polygon cuts off some small triangular pieces of the original rectangles and that it also includes some small triangles not formerly included. But is obvious that triangle A = triangle ; triangle B = triangle B and so on.

Sometimes the Frequency Polygon is extended at each end to join the X-axis (indicating zero frequency ) at the midpoint of the next possible class. This procedure results in having the same area under the frequency polygon as is included in the histogram. However, in our illustration, this procedure results in extending the Polygon beyond zero point on the x-axis and this is to be meaningless. In any event the extensions will suggest to the reader that items accrued beyond the limits of the observed data. Hence, except for special purposes, it is better not to extend the frequency polygon to the X-axis.

Above we have explained the method of drawing the frequency polygon using histogram However, it is not always essential first to draw histogram in order to obtain, frequency polygon Frequency polygon may also be constructed as explained below.

(i) Mid-values of the classes of the given frequency distribution are obtained first.

(ii) These mid-values are marked on the X-axis taking a suitable scale.

(iii) The frequencies (in the case of equal class intervals ) or frequency densities (in the case of unequal class intervals) are marked on the Y-axis.

(iv) Above each mid-value and opposite to the corresponding frequency (or frequency density) we mark a point. All such points plotted are joined by means of straight lines and the Frequency Polygon is obtained.

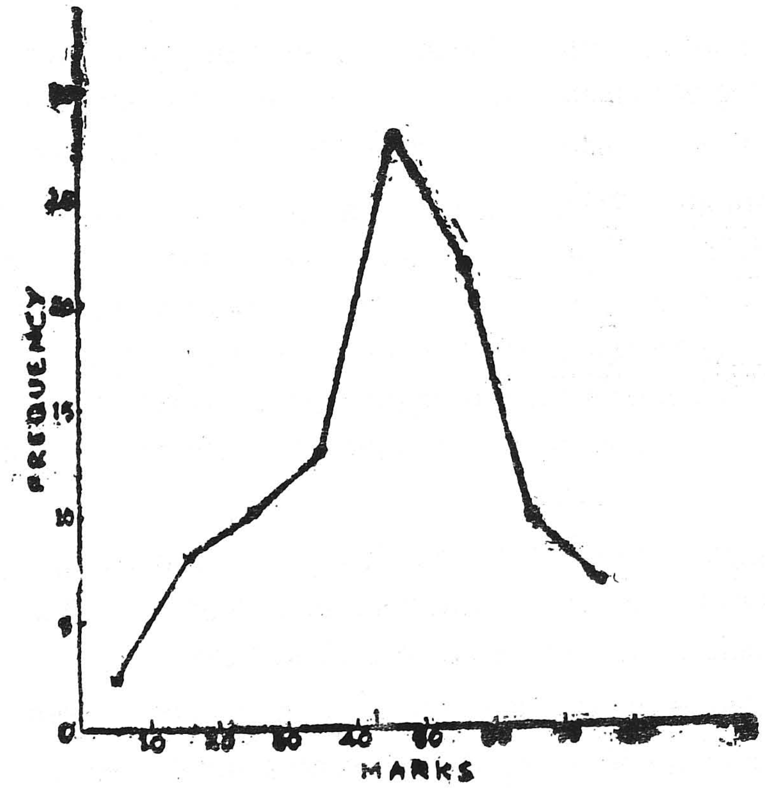
Consider the following example.

### Illustration - 6

Consider the same data given under Illustration - I and draw the frequency Polygon.

| Class | Mid-value | Frequency |
|---|---|---|
| 0-10 | 5 | 2 |
| 10-20 | 15 | 8 |
| 20-30 | 25 | 10 |
| 30-40 | 35 | 13 |
| 40-50 | 45 | 28 |
| 50-60 | 55 | 22 |
| 60-70 | 65 | 10 |
| 70-80 | 75 | 7 |

Now we mark the mid-values on the X-axis and the frequencies on the Y-axis and get the frequency polygon as shown in the following diagram.



## Histograms Versus Frequency Polygons

This Histogram has the following advantages over the Frequency Polygon :

(i) Each columns in a histogram represents the exact number of items in a class. But the areas over any class in the frequency polygon does not represent the exact number of items in that class. Because, the area cut off is not exactly equal to the area added to a column when a frequency polygon is drawn. Also the column corresponding to the highest frequency loses without gaining anything.

(ii) The individual classes stand out more clearly in the case of a Histogram than in a Frequency Polygon.

(iii) In the case of unequal class intervals, the Histogram makes it easier for the reader to note the changing width. The changing width will not be apparent in the case of Frequency Polygon.

(iv) Histogram can be drawn to represent a discrete frequency distribution also. Here, separate bars are drawn to emphasis gaps in a discrete frequency distribution. But, Frequency polygon cannot be drawn to represent a discrete frequency distribution.

## Frequency Polygons have the following advantages over Histograms.

(i) A frequency polygon looks more simple than a histogram because the former has fewer lines compared to the latter.

(ii) To compare two distributions we cannot draw their histograms on the same graph paper. Because, they will make numerous confusing intersections. But, two frequency polygons drawn on the same graph paper will make very few intersections and hence they can be compared easily.

Histograms are usually preferable when classes are few and frequency polygons when classes are numerous. However, either type of graph can ordinarily be used.

Both histograms and frequency polygons have an advantage that is characteristic of all diagram an graphs they summarize or generalize data and so provide a fast and relatively easy method of presenting experience. A manufacturer or a retailer from his sales records, can prepare frequency graphs showing the sizes os shoes, clothing add other merchandise of his customers and use them a guides in future purchasing and inventory control.

## Comparison of two frequency distributions

It is often desirable to compare graphs or two frequency distributions on the same chart. For such comparisons only frequency polygons should be used. Because, if we draw two histograms on the same chart their columns will overlap, thus blurring the picture with confusing intersections.

For the graphic comparison of two distributions using frequency polygons.

(i) The two distributions should be in the same units of measurement.

(ii) The two distributions should have the same class intervals; if the class intervals are different, frequency densities per unit must be used instead of frequencies.

(iii) The two distributions should have nearly equal number of items. If the two distributions differ appreciably in regard to the total number of items involved, then percentage frequencies are calculated expressing the total frequency of each distribution as 100 percent.

If two distributions are expressed in different units (one in rupees and the other in inches) then a direct graphic comparison is not possible.

## Illustration - 7

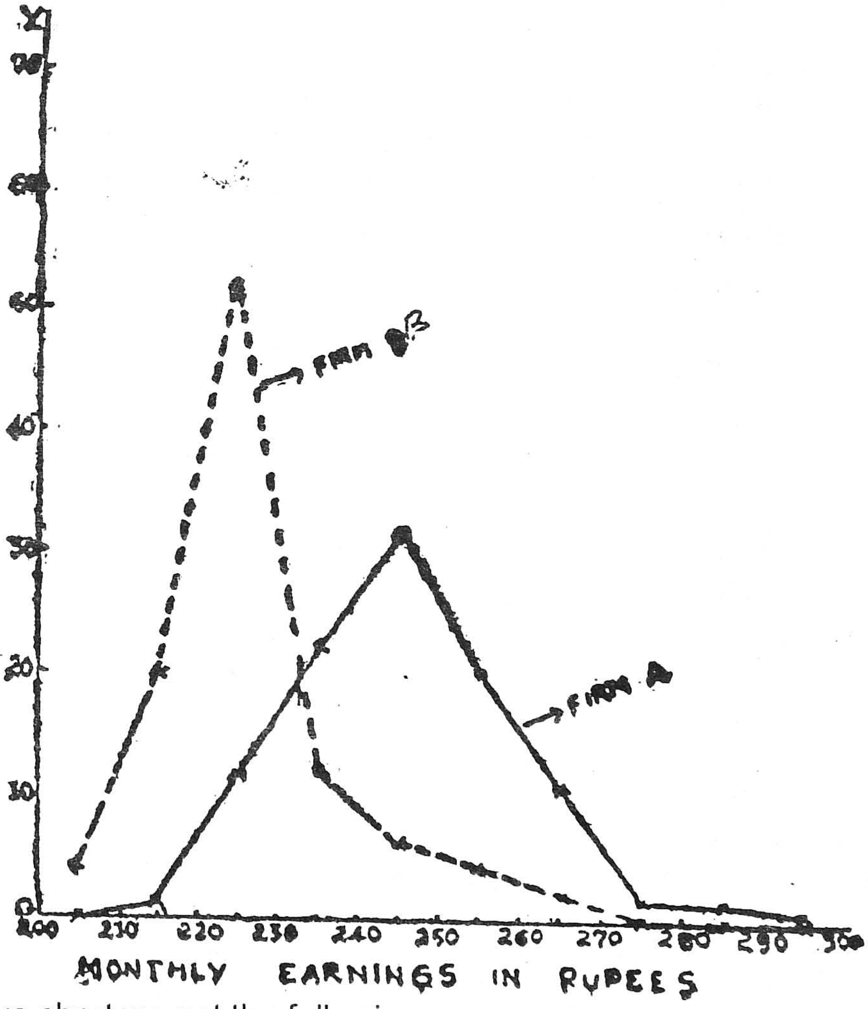
Monthly earnings of workers of two firms A and B are as follows. Draw the frequency polygon and compare the two distributions.

| Monthly earning in Rs. | Number of workers | |
|---|---|---|
| | Firm A | Firm B |
| 200-210 | 0 | 2 |
| 210-220 | 2 | 10 |
| 220-230 | 23 | 26 |
| 230-240 | 44 | 6 |
| 240-250 | 63 | 3 |
| 250-260 | 40 | 2 |
| 260-270 | 21 | 1 |
| 270-280 | 3 | 0 |
| 280-290 | 3 | 0 |
| 290-300 | 1 | 0 |
| Total | 200 | 50 |

The above Table consists of two frequency distributions with the same classes expressed in the same units, namely rupees. So, we can make a graphic comparison of the two distributions using Frequency Polygons. However, the two distributions vary widely in regard to their total number of items - total number of workers in firm A is four times that of in firm B. Hence, we first compute the percentage frequencies equating the total frequency of each distribution to 100 and then proceed with the drawing of Frequency Polygons. We have given the percentage frequencies in the following Table.

| Monthly earning in Rs. | Percentage of workers | |
|---|---|---|
| | Firm A | Firm B |
| 200-210 | 0 | 4 |
| 210-200 | 1 | 20 |
| 220-230 | 11.5 | 52 |
| 230-240 | 22 | 12 |
| 240-250 | 37.5 | 6 |
| 250-260 | 20 | 4 |
| 260-270 | 10.5 | 2 |
| 270-280 | 1.5 | 0 |
| 280-290 | 1.5 | 0 |
| 290-300 | 1.5 | 0 |
| Total | 100 | 100 |

**MONTHLY EARNINGS OF WORKINGS OF TWO FIRMS A AND B**




From the above chart we get the following :

(i) Workers in Firm A earn more than workers in Firm B for the most part.

(ii) The most usual earning rates are in the Rs.220. Rs. -230 bracket * for Firm B as compared with Rs. 240 Rs. -250 bracket for Firm A.

(iii) There is a much greater concentration of earning in Firm B than in Firm A in these respective brackets. This is shown by the relative heights of the polygons in the brackets.

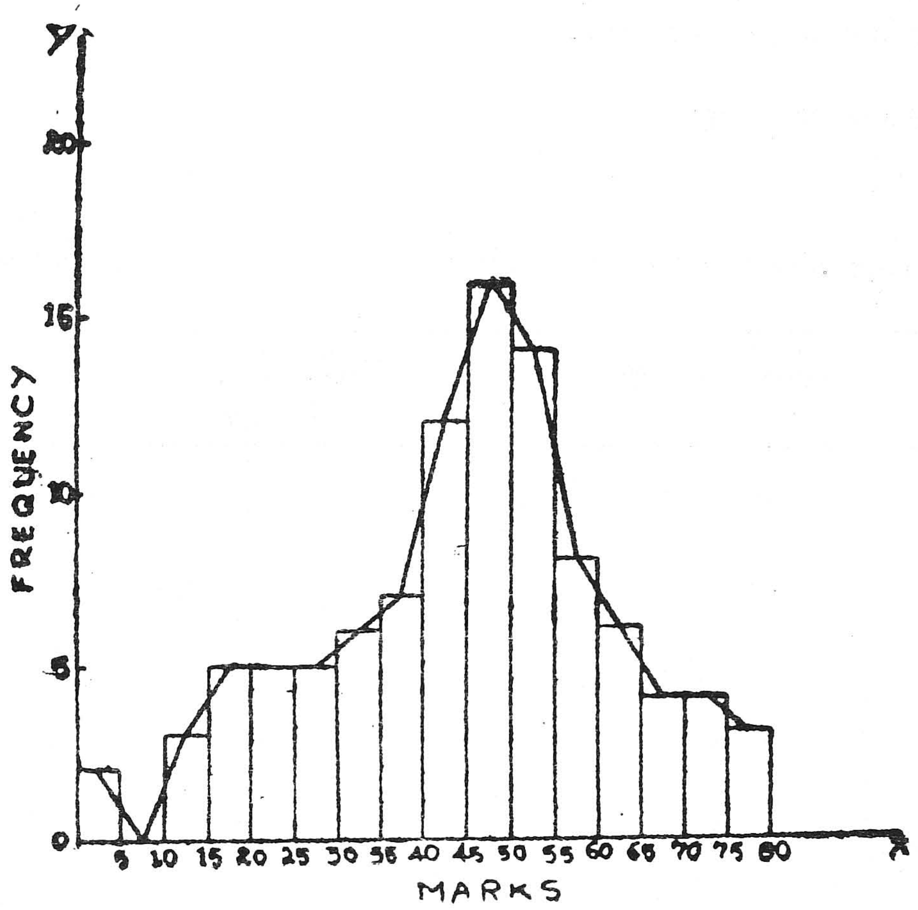
## FREQUENCY CURVE

The Table given under illustration I shows the distribution of marks obtained by 100 students, arranged according to class intervals of 10 marks each. We have the corresponding Histogram and Frequency Polygon in the same chart under illustration -5. The marks obtained by the students range from 0 to 80. If we could make our class intervals smaller and smaller the columns the Histogram would become narrower and narrower. Also, the number of students belonging to each class would also become smaller and smaller if we reduce the width of the class intervals from 10 marks to 5 marks. Suppose the table given under illustration - I takes the following form.

| Marks | No. of students |
|---|---|
| 0-5 | 2 |
| 5-10 | 0 |
| 10-15 | 3 |
| 15-20 | 5 |
| 20-25 | 5 |
| 25-30 | 5 |
| 30-35 | 6 |
| 35-40 | 7 |
| 40-45 | 12 |
| 45-50 | 16 |
| 50-55 | 14 |
| 50-60 | 8 |
| 60-65 | 6 |
| 65-70 | 4 |
| 70-75 | 4 |
| 75-80 | 3 |
| Total | 100 |

* The interval is which the peak of the frequency polygon lies represents the most usual earning rates bracket.

The corresponding histogram and frequency polygon draw on the same graph are as follows.




If we reduce the class intervals without increasing the total number of items, it might happen that a particular bar in the Histogram might become very small in length or even might disappear. In our example, the second bar has disappeared because, the frequency of the second class, namely 5-10 has become zero. But if we can get the marks obtained by larger number of students, we can make the bars narrower without making them disappear altogether when we reduce the width of the class intervals and increase the number of items simultaneously, the frequency polygon takes more and more an appearance of a smooth curve. Thus, the effect of increasing the total frequency and decreasing the class width would smooth out (ie.) remove the sharp comers) the frequency polygon and it looks like a curve. The curve obtained from smoothing a Frequency Polygon is called 'Frequency curve' or 'smoothed Frequency curve'.

Sometimes it may not be possible to study large number of items and smaller size of class intervals. In such a case, the smoothened frequency curve is obtained as follows :

1) The histogram of the given frequency distribution drawn on the graph paper.

2) The midpoints of the tops of the rectangles in the Histogram are marked.

3) Through these midpoints we draw a curve by our free hand in such a way that the areas gained and lost by each column are equal. *Hence, the frequency curve must always overlap the height column of the Histogram.

Consider the following examples

## Illustration - 8

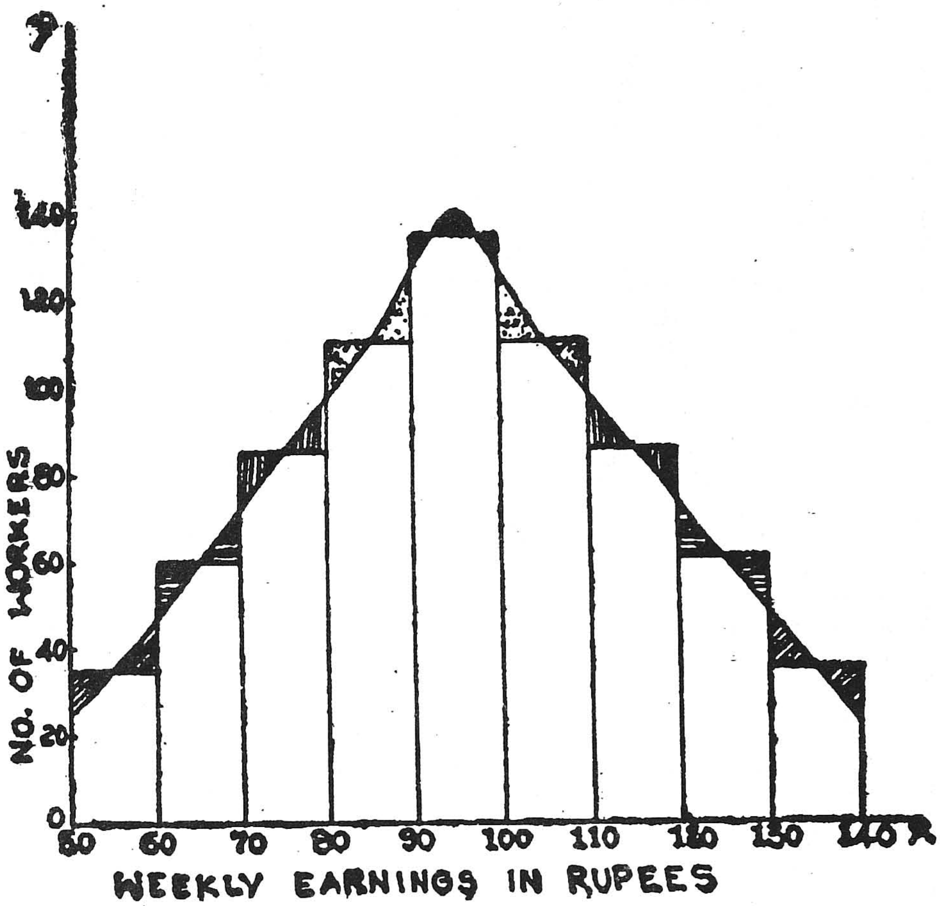
Weekly earnings of the workers of a factory are as follows. Draw the frequency curve :

| Weekly earnings | No. of workers in Rs. |
|---|---|
| 50-60 | 35 |
| 60-70 | 60 |
| 70-80 | 85 |
| 80-90 | 110 |
| 90-100 | 135 |
| 100-110 | 110 |
| 110-120 | 88 |
| 120-130 | 60 |
| 130-140 | 35 |
| Total | 715 |

$ A curve which does not contain sudden turns or on other words sharp corners is called a smooth curve.

* Here note the difference between the frequency polygon and the frequency curve. While drawing the polygon also, each column of the histogram losses a triangular area and gains another triangular area. If we take, a single column, the area gained and lost will not be equal. But, the area lost by one column will be exactly equal to the area lost by the immediately succeeding column. In the case of frequency curve also, each column of the histogram loses certain area and gains certain area. Here, the area gained and lost must be equal for each individual column.

The Frequency curve is as follows :




In the above diagram, we have given the same shading to the area cut off of and the area added to compensate it. The same shading is used to indicate that the two areas are equal.

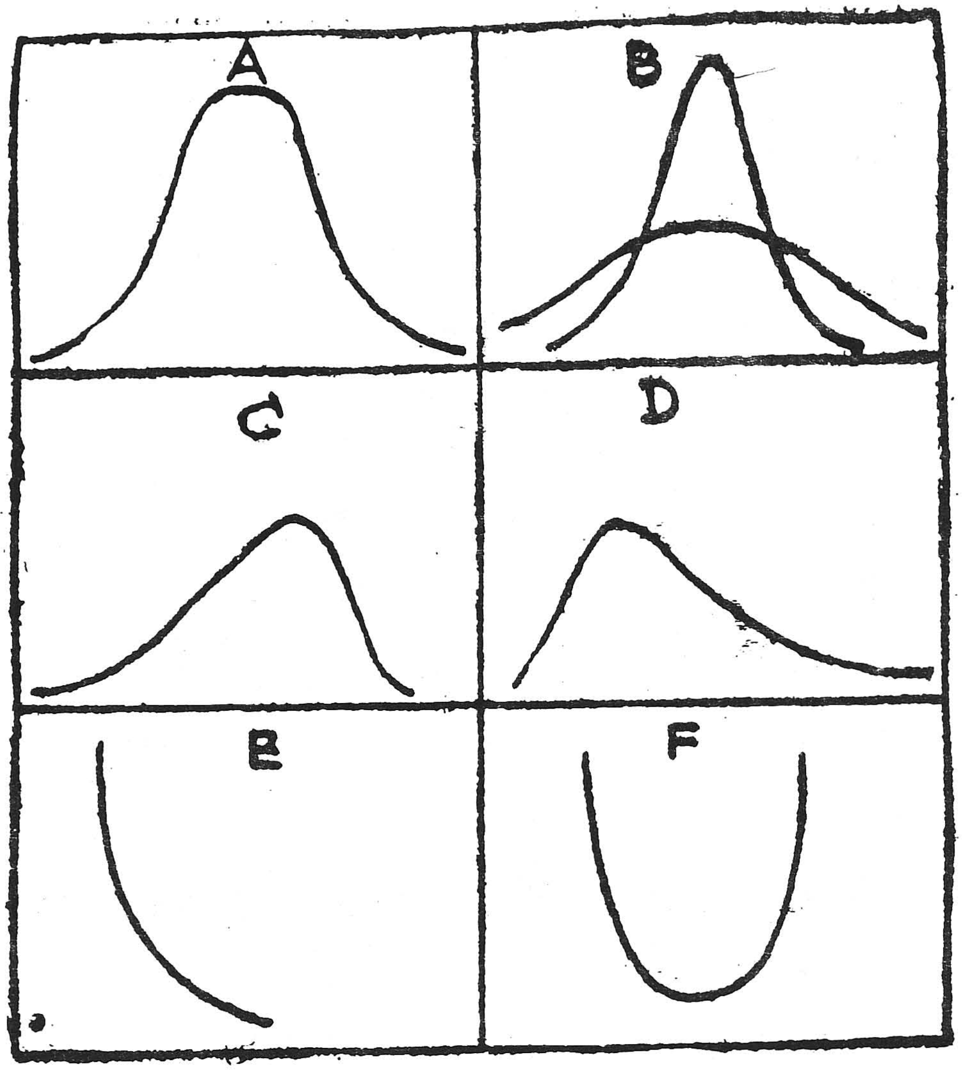
As the frequency curve is drawn in such a way that the areas lost and gained by each column of the histogram are equal, the area under the frequency curve and the are of the Histogram, will be equal. Thus, the total area under the frequency curve will exactly represent the total frequency.

Frequency curves are drawn only for frequency distributions based on samples. Through the process of smoothing, accidental irregularities present in the data are eliminated. Thus, the purpose of drawing a smoothed frequency curve is to indicate the normal tendency within this distribution and within each class. That is, the Frequency Curve is designed to indicate the nature of what the distribution of the whole universe would be if all of the values were plotted instead of a sample grouped by classes.

If the given set of data begins and ends with a zero or a very small number of occurrences, the frequency curve is drawn in such a way that it begins and ends with the X-axis. Otherwise, we need not extend the frequency curve on both ends to meet the X-axis.

## Types of Frequency distribution

Different types of frequency distributions are usually described in terms of smooth curves. These smooth curves usually assume anyone of the following six shapes :




The curves given in diagrams A and B are symmetrical curves. The curve given in A looks like a bell; it is moderately peaked and it is called normal curve.

The normal curve is a picture of the variations due to chance. In practice, business and economic data are seldom exactly normal. However, many distributions in these fields are close enough to the shape of normal curve and hence, its characteristics are usual in their analysis. In addition, normal curve is indispensable in the study of reliability of averages obtained from samples.

Curves given in B are symmetrical like normal curve but one is flatter and the other more sharply peaked. These shapes result from the influence of extraneous factors which offset the chance variations.

Curves given in C and D have a long tail on either side. Such curves are called 'asymmetrical' or 'skewed' curved. Curve has a long tail on the left and its called 'negatively skewed' curve. This type of curve commonly result from a distribution having a fixed upper limit but a more remote lower limit. Only rarely we come across negatively skewed curves.

---

@ Through the apex of the curve, if we draw a vertical line and fold the graph paper, the left hand portion of the curve will exactly coincide with the right hand portion if it is a symmetrical curve.

Curve D has a long tail to the right and it is called 'positively, skewed' curve. This is the most common type of curve encountered in most business and economic data. Distributions of personal earnings, commodity prices or assets of companies usually tend to have a shape of positively skewed curve.

The curve given in diagram E is reverse J-shaped curve. Distributions of sizes of stores, cities, income tax payments etc., usually approximate to the reverse J-shaped curve. This curve denotes that the smallest sizes are numerous and the number of sizes drops off sharply at first and then more gradually increases as the size increases.

The last curve given in diagram F is U- shaped. This curve denotes that the smallest sizes are great in number then of sizes drops off sharply, it reaches a minimum and then the number of sizes increases again. U-shaped curves are less common in practice.

## OGIVES

### Cumulative Frequency Graphs or Ogives

Sometimes especially in the case of distributions having open intervals at their ends it is useful to cumulate the frequencies of the distribution. Graphic representation of a cumulative frequency distribution is called a cumulative frequency graph' or an 'ogive' (pronounced ogive).

We have already seen that there are two types of cumulative frequency distributions, namely, less than cumulative frequency distribution and more than cumulative frequency distribution. The graph representing the former is called 'less than ogive' while the graph representing the latter is called a 'more than ogive'.

To draw an ogive, the true class limits are marked on the x-axis and the cumulative frequencies are marked on the Y -axis.

Against the true upper limits of the class intervals points are plotted corresponding to the respective less than cumulative frequencies. The line joining the points plotted as above gives the 'less than ogive'.

Against the true lower limit of the class intervals, points are plotted corresponding to the respective more than cumulative frequencies. These points are joined by a line and the graph so obtained is the 'more than Ogive'.

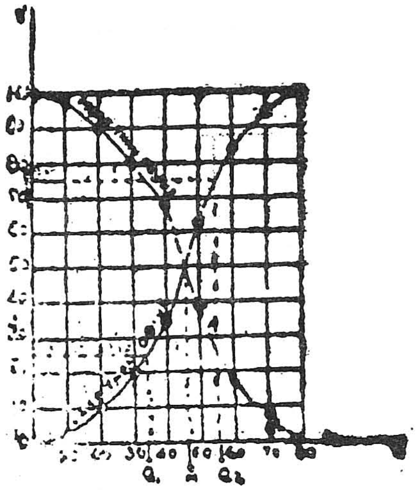
### Illustration - 9

Cumulative frequency distributions for the same Table given under illustration 1 are given below :

Draw the less than and more than ogive with the use of the data.

| Marks less than | Frequency | Marks greater than | Frequency |
|---|---|---|---|
| 0 | 0 | 0 | 100 |
| 10 | 2 | 10 | 98 |
| 20 | 10 | 20 | 90 |
| 30 | 20 | 30 | 80 |
| 40 | 33 | 40 | 67 |
| 50 | 61 | 50 | 39 |
| 60 | 83 | 60 | 17 |
| 70 | 93 | 70 | 7 |
| 80 | 100 | 80 | 0 |

'Less than' and 'more than' Ogive



If we are given a simple frequency distribution and are asked to draw the ogives, first of all we must form the less than and more than cumulative frequency distributions. While forming the cumulative frequency distributions, we must see whether the given classes are true classes or not. If not, we must convert the given classes into true classes and then proceed to form the cumulative frequency distributions.

## Illustration - 10

Draw the less than and more than ogives for the following frequency distribution.

| Class Intervals | Frequency |
|---|---|
| 0-9 | 8 |
| 10-19 | 12 |
| 20-29 | 17 |
| 30-39 | 15 |
| 40-49 | 20 |
| 50-59 | 10 |
| 60-69 | 5 |

In the above Table, the class intervals are not true class intervals. So, first we convert them into true class intervals and then the less than and more than cumulative frequencyh distributions.

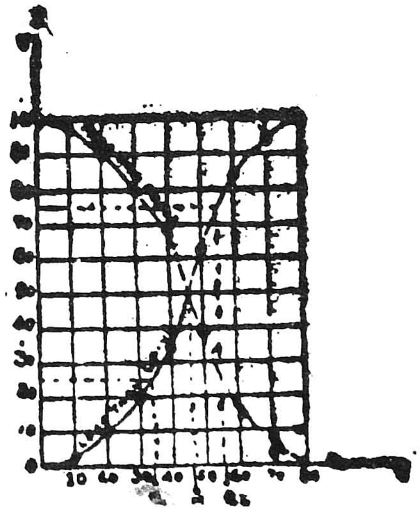
| True class intervals | Frequency |
|---|---|
| .5 - 9.5 | 8 |
| 9.5 - 19.5 | 12 |
| 19.5 -29.5 | 17 |
| 29.5 - 39.5 | 15 |
| 39.5 - 49.5 | 20 |
| 49.5 - 59.5 | 10 |
| 59.5 - 69.5 | 5 |

We have given below the less than and more than cumulative frequency distributions :

| Less than | Frequency | More than | Frequency |
|---|---|---|---|
| -.5 | 0 | -.5 | 87 |
| 9.5 | 8 | 9.5 | 79 |
| 19.5 | 20 | 19.5 | 67 |
| 29.5 | 37 | 29.5 | 50 |
| 39.5 | 52 | 39.5 | 35 |
| 49.5 | 72 | 49.5 | 15 |
| 59.5 | 82 | 59.5 | 5 |
| 69.5 | 87 | 69.5 | 0 |

'Less than and More than' Ogives are given in the following figure.

**OGIVES**



## Advantages of Ogives over Histograms and Frequency Polygons

1. Presence of open class intervals does not, in any way affect the drawing of ogives.

2. No adjustment need be made in the case of a distribution having unequal class intervals.

3. Even when the class intervals of two or more distributions are not uniform they can be compared with each other with the help of ogives. However, the given distributions must be in the same units and must contain nearly the same number of items. If the number of items vary widely between the given distributions percentage cumulative frequencies must be calculated and used to draw the Ogives and than only the comparison must be made.

## Usefulness of Ogives

1. Ogives are often used to present data of wages and of hours of work.

   With reference to wages, the less than ogive enables to ascertain how many of a group receive less than a subsistence level standard level or comfort level. Similarly, the more than ogive enables us to ascertain the number receiving a subsistence level or more a standard level or more and a comfort level or more. It is also possible to ascertain what wage (the lowest or the highest) is paid and the number of workers or the percentage of workers receiving it.

   With respect to hours of work, we can see quickly the number of people working unusually long or short hours.

2. Generally, there is less overlapping when comparing several ogives in the same graph than when comparing several Frequency Polygons in this manner. Therefore, to compare two or more frequency distributions Ogives are much useful.

3. Ogives are also useful to find out certain statistical measures like median, quartiles, declines and percentiles.

## Model Question

Describe the methods of presentation of statistical information like histograms, frequency polygon, frequency curve and ogives.

## LESSON -10
# SAMPLING AND ITS DISTRIBUTION

When the researcher has completed the task of formulating the problem and developed a research design he has to decide whether the enquiry should be conducted from all the members of the society under investigation or only a percentage of the concerned group. If the data is collected from all the members of the concerned population it is known as census survey. On the contrary, if the data is collected from only a fraction of the members of the particular society concerned it is known as the sample survey. The researcher on the basis of the nature of the problem has to decide whether he has to conduct a sample survey or a census survey.

## Types of population

Depending upon the number of items populations may be classified into two types namely, finite and infinite populations.

Consider the population of electric bulbs produced by a firm. Total number of bulbs produced by the firm will definitely be a known finite number. Here, the population is called a finite population. On the other hand, consider the population of stars in the sky. We cannot find out the total number of stars in the sky because they are infinitely large in number. Here, the population is called infinite population.

Depending upon the nature of the items of the population they may also be classified as hypothetical and existent universes. Hypothetical universe is one which does not consist of concrete objects. If we toss a coin the head or tail may turn upwards. By throwing the coin large number of times and by recording the result of each throw we can construct a population. The result of each throw is an item of this population and these items are not concrete objects. Such a population is called a hypothetical population; on the other hand, if we consider the population of electric bulbs

Another type of distinction may also be made between populations. Suppose a grocer makes a survey of customers opinions about a new brand of soap supplied by him. Here all his customers constitute the population. Here the population consists of human beings. Hence, this population is called human population. On the other hand, suppose a survey of the life times of the electric bulbs produced by a firm is made. Here the electric bulbs constitute the population. This is called inanimate population studies related to human populations are typical of problems encountered in market research and demography, whereas studies of inanimate populations are typical of industrial quality control and accounting. The population may also consist of animals as in biological studies. Such populations are called animal populations.

---

* Ordinarily, the word 'population' is used to denote the total number of inhabitants of a county, town, etc. But, in Statistics it is used in a technical sense to refer to the totality of items from which the sample has been drawn. Thus, in Statistics, we talk of population of prices, population of rents, population of electric bulbs, population of men's heights etc. produced by a firm or the population of number of children per family in a city. These populations consist of concrete objects as their items. Such populations are called existent populations.

There are some basic terms to be known by the researcher while discussing sampling. They are population, sampling unit, frame and sampling design.

## 1. Population

The term population refers to any finite or infinite collection of individuals eg. population of industrial labourers in an urban area.

## 2. Sampling unit

One of the units into which one of the aggregates is divided for the purpose of sampling each unit being recorded as individual and indivisible. When the selection is made eg. households, persons etc.

## 3. Frame

A list, map or other specification of the unit, which constitute the available information relating to the population designated for a particular sampling scheme. Eg. Telephone Directory.

## 4. Sampling design

It is a definite plan for obtaining a sample from the sample frame. It refers to the technique or procedure the researcher would adopt in selecting the items for the sample.

## Sample

A sample is that part of the universe which we select for the purpose of investigation. A sample should exhibit the characteristics of the universe; it should be a "microcosm", a word which literally means "small universe".

Everyday in the ordinary business of life we make judgements and take action based on samples. When we go to a shop we examine a handful of rice to find the quality of rice in the whole bag . The housewife examines a ladle of rice to know if the rice is well cooked. A doctor examines a few drops of blood to draw conclusion about the blood constitution of the whole body. The teacher asks a student a few question to determine his grasp of an entire subject. A business man places order for commodity by examining only a few units of the same. Thus, there is hardly any field in which the sample technique is not used.

## Census versus sample method

Information on a population may be collected in two ways. Either every unit in the population is enumerated or enumeration is limited only to a part (ie. a sample) selected from the population. The former method of enumeration is called complete enumeration or census while the latter is called sample enumeration or sample survey.

Complete enumeration is always much more extensive and time consuming. For instance, in establishing a brand preference for soap it would cost too much money and consume too much time to ask every person in India, even if complete enumeration is done once. It may prove too costly in money and time if investigation has to be repeated, as it must be in periodic market

research to keep check on consumer preference. Moreover, complete enumeration may take so much time that the findings are of no use by the time they become available.

The economy of cost and time is not the only consideration for the use of a particular method of enumeration; the most important point is whether the accuracy of the results would be adequate for the end in view. It is a curious fact that the results from a carefully planned and well-executed sample survey are expected to be more accurate than those from a complete census that can be taken. A complete census ordinarily requires a huge and unwieldy organisation and therefore many types of errors creep in which cannot be controlled adequately. In a sample survey the volume of work is much reduced and hence it becomes possible to employ a person of higher calibre, train them suitably and supervise their work adequately, Also more detailed information may be obtained under sample method. There may be errors in the sampling process. But the errors are usually much less than the errors due to the inaccuracy and incompleteness of the large scale survey.

In a properly designed sample survey it is possible to make a valid estimate of the margin of error and hence decide whether the results are sufficiently accurate. It is also possible to devise ways of reducing the errors due to sampling. A complete census does not reveal by itself the margin of uncertainty to which it is subject to .

However, there is not always a choice of one versus the other. In cases where destructive tests are to be conducted or where the population consists of very large number of items, only the sample method is practicable. For example, suppose we want to test the strength of the beams supplied by a firm. Here we cannot adopt the census enumeration which implies loading of each and every beam till it breaks. Here, we have to select a sample of few beams test their strength and come to a conclusion regarding the strength of the whole lot supplied by the firm. Similarly, to test the quality of rice in the whole bag, we cannot test the quality of each and every rice in the bag. Here also only the sample method is a practicable one.

If information about every item of the population is required, only a complete census can deliver the desired information. For example, to find out the total number of adults, literates, etc. no sample survey of a reasonable size can give use the required information. In these cases we have to resort to census method only.

## Advantages of Sampling

Sampling is used very frequently in marketing research because it offers some benefits over taking a census.

### 1) Economy in expenditure

Survey of small number of respondents requires less money. For instance, in establishing a brand preference for soap, it would cost much time and money to ask every person in the country. If data are collected from a small sample of the universe, the expenditure would be less and hence economical.

## 2) Saving of time

Survey of smaller number of units requires much less time than census method. The data can be collected and summarised more quickly by studying sample than studying the universe. This may be a vital consideration when the information is urgently needed or the result of the study has to be declared quite early.

## 3) Greater accuracy

The results drawn by sampling technique are more reliable than the results from census method. The smaller the study, the more likely we are to obtain more highly skilled people for each stage of the research process. If the sample has been properly selected, the results are within a very close range of a accuracy.

## 4) Detailed study

When the number of units is large, detailed study is not possible. The smaller number permits a more minute observation and detailed study.

## 5) Practicability

When the universe is extremely larger, sampling is the only practical method of analysis. It would be a tremendous job to interview every person in a town to know about his preference for tea or coffee to know about is population. The only practical procedure is to study a samples.

## Law of Statistical regularity and Law of Inertia of large numbers

The whole theory of sampling is based upon two important principles which are also referred to as the 'laws of sampling'. However they are not laws in the strict sense of the term; they are rather principles and tendencies which appear to operate universally. The two principles of sampling are as follows :

(i) Law of Statistical Regularity and

(ii) Law of Inertia of Large Numbers

## Law of Statistical Regularity

Everything in nature and life occurs with a surprising regularity. Even accidents which are comparatively rare events do not happen haphazardly as people think. If 50 students are selected at random from a class containing 400 students, the average weight measurement for them will differ but little from that of the entire class. If 100 sugarcanes are taken at random from a field where sugarcanes are grown and the average height of this sample is calculated it will not differ by more than a small fraction from the average height of all the sugarcanes in the entire field. These are manifestations of the Law of Statistical Regularity.

## This Law may be stated as follows

If a sample is selected at random from a population, the characteristics of this sample will differ but by a very small percentage from those of the entire universe.

The important condition for the principle to function is that the items for the sample should be selected at random. By random selection, we mean a selection where each and every item in the population has an equal chance of being included in the sample. That is, there should be no specific reason for the inclusion of a particular item in the sample.

By reliance upon the principle of statistical regularity it is possible to estimate fairly accurately the characteristics of the entire universe by studying only a small part (ie. a sample) of the universe. It is also clear that this principle makes possible a material reduction of the work necessary before accurate conclusions may be made concerning any large group of quantitative data. Thus, if one is desirous of studying the changes in the earnings of factory workers in India, it would not be necessary to include all the workers belonging to all the factories. If the earnings records of say, 25 percent of the workers in each of some 25 percent of factories were studied almost as good a view of the changes in the earnings of factory workers could be obtained as if all the factory workers were included in the analysis.

## Law of Inertia of Large Numbers

A natural consequence of the law of statistical regularity is the "Law of Inertia of Large Numbers'. There may be many changes in the individual items. But, when we study the entire lot together the change will be slight. That is large numbers have greater stability, steadiness and consistency than that possessed by small ones. It is based on the principle that the changes in the individual components of a population are not cumulative but compensating. That is, while some items move in one direction others move in exactly opposite direction. This is said to be inertia of large numbers meaning thereby that large numbers together change slowly.

It may be that in one year some have become paupers and a few have become rich in a particular city. But the sum total of the wealth of the entire city would have changed only a little. The rice crop in one or two districts may have failed completely due to late rains or floods. But the total production of rice in South India will show only very little change from year to year. In a bank with a large number of branches, the number of customers having accounts with balances below Rs.100 may vary greatly from week to week among the different branches. That is, some branches may have a smaller number than the preceding week while the other branches may have a large number. However for the bank as a whole, the number of accounts less than Rs. 100 may not change greatly from one week to the next. To use another common example, the fire losses in a particular city or in a small group of cities selected at random may be considerably large per month this year than last. But the total loss due to fire in the country as a whole shows little change from year to year. This is the reason why fire insurance companies are able to base their rates so low and yet work on a profit.

An important application of the principle of inertia of large numbers is in the field of life insurance. The highly remunerative business of life insurance is dependent upon estimating how many people in a given age group will die each year. Experience tables of insurance companies cannot show who probably will die; but they do reveal how many probably will die. Some x may die in his 65th year of age, some y may die in 50th year and so on. Thus the life spans of individuals are quite erratic. Yet a probable given number of several thousand will die at a particular age. This is some what a stable characteristic.

## The sampling Process

An over view of the steps in selecting a sample.

Step :1. Define the population

2. Identity the sampling frame
3. Decide on sample size
4. Select a sampling plan
5. Select the sample.

### 1. Define the population

Population may be defined as the larger group which the researcher wants to know about studying one or more of its samples. The population must be defined in terms of elements, sampling unit, extent and time. For example, if we were to conduct a survey on consumption of tea in Tamilnadu, the population may be defined as follows ;

| | | |
|---|---|---|
| Element | : | House wives |
| Sample unit | : | Households |
| Extent | : | Tamilnadu |
| Time | : | 3 Months |

### 2) Identify sampling frame

In marketing research, most of the frames are from census reports, electoral register, telephone directory, list of members of trade and industry association, list of members of professional bodies, list of dwelling units maintained by local bodies etc. If the frame is incomplete or otherwise detective, sampling will not be able the overcome the shortcomings.

### 3) Decide on a sample size

Here, on has to decide how many elements of target population are to be included in the sample.

### 4) Select a sampling plan

The operational procedure for the selection of the sampling are selected.

### 5) Types of sampling designs

The samples designs are basically of two types viz., probability sampling and non-profitability sampling.

In probability sampling each element of the population has a chance of being selected for the sample. In non-probability sampling, the selection of a population elements for a sample in influenced by personal judgement.

## Probability sampling

1. Simple random sampling
2. Stratified random sampling
3. Systematic sampling
4. Cluster sampling
5. Multistage sampling
6. Area sampling
7. Sequential sampling

## Non- probability sampling

1. Convenience sampling
2. Judgement sampling
3. Quota sampling

## Sampling Methods

The process of selecting a sample is called 'sampling'. There are different methods of selection of samples. The choice of a suitable method depends upon (i) the nature of the problem under consideration (ii) the nature of the population from which sample is to be drawn and (iii) the uses for which the results obtained from the sample are to be put.

Various methods of designing samples can be broadly classified into two types. They are :

1. Random or probability sampling (probability sampling)
2. Purposive or deliberate or judgement sampling (Non-probability sampling)

## Random Sampling

Random sample is one where the individual items constituting the sample are selected at random. "Random" is not used here in the sense of "haphazard" or "hit or miss". By random we mean that the selection has been made in such a manner that each item of the population has equal or known opportunity of being included in the sample. Chance and only chance determines which items are to be included in the sample chosen.

Random samples may again be classified into two types known as (i) unrestricted or simple random samples and (ii) restricted random samples.

Unrestricted random samples are usually obtained by using any one of the following two methods.

(i) Lottery method.

(ii) Random numbers method.

## Lottery Method

The process of drawing by lot is also known as simple random sampling. In this method various items are represented by small chits of paper of the same size which are folded in the same manner and mixed together. Form this the required numbers are picked out by a blind-folded person. For example, if we want a sample having 10 items then 10 chits are picket out. The items represented by the 10 chits picked out from the required sample.

The method becomes tedious when the size of the population or the size of the sample to be selected is large.

## Random numbers method

By selecting 4,600 digits from census returns, Tippett, combined them by fours to form 10400 four digit numbers. They are given in tables known as 'Random sampling numbers'. The first twenty numbers given in Tippett's table are as follows.

| | | | |
|---|---|---|---|
| 2952 | 6641 | 3992 | 9792 |
| 4167 | 9524 | 1645 | 1326 |
| 2370 | 7483 | 3498 | 2762 |
| 0560 | 5246 | 1112 | 6107 |
| 2754 | 9143 | 1405 | 9025 |

The technique of selecting random sample with the help of these numbers is very simple. Suppose we have to select a sample of 15 items from a population of 4000 items. First we shall number the various items of population from 1 to 4000. We shall open any page of the Tippetts Table and select the first 15 numbers which are less than 4000. The items having these numbers form the required random sample.

Though Tippett's numbers are of four digits in many cases they are less than 1000 or 100 or 10. Consider the Tippett's numbers 0003, 0560 0104. The number 0008 is simply 8 and it is less than 10. The other two numbers are less than 1006. Hence even if the population is small say of 100 units only and if we have to take a sample of 10, we can use Tippett numbers and get the sample.

Tippett numbers have tested several times with known population and the randomness of these numbers is almost beyond question. Similar tables have been devised by Fisher, Kendall and Mahalanobis. They are also very popular and have given correct result in large number of investigations.

So far we have seen the methods of getting unrestricted random samples. Unrestricted random samples usually result in greater cost or less degree of precision. Hence, in order to increase efficiency a restricted random sample is used. A restricted random sample may be obtained in anyone of the following three methods.

(i) Stratified sampling

(ii) Cluster Sampling

(iii) Systematic sampling

## Stratified sampling

Stratification means dividing an entire universe into groups. These groups or subuniverses are called strata. When the universe consists of heterogeneous items, it is divided into various homogeneous groups so that the efficiency of sampling is increased.

Suppose we want to conduct an enquiry into the standard of living of people in a city. We know that the standard of living is determined by the income level of people and there will be marked difference in the standard of living of high income and low income people. So if we take the universe as a whole and draw a sample by unrestricted random sampling method it may be a true representative of the universe. More reliable results may be obtained if we divide the population into high, medium and low income groups and then draw a sample.

The individual problem under investigation determines the characteristics that should be used for stratification. In practice, geographical, sociological and economic characteristics are often used for the grouping of universe.

After the stratification is over, the next problem is to decide upon the number of items to be drawn from each stratum. The procedure employed in determining the number of items to be drawn from each stratum is termed allocation. One simple procedure is to draw equal number of items from each stratum. For instance, to find out the standard of living of people in a city suppose we want to select a sample of size 30. To get this sample 10 people may be selected from low income group, 10 from medium income group and the remaining 10 from high income group. Here irrespective of the total number of items belonging to each stratum, we have drawn equal number of items from each stratum. In this procedure we lose many of the advantages of stratification. There are three more suitable methods of allocation.

***(i) Allocation proportional to sizes of strata :*** In this method, items are drawn in proportion to the sizes of each stratum. Suppose, in a city 10,000 people belong to low income group, 4000 people belong to medium income group and 1000 people to high income group.

Here, items are drawn from each stratum in such a way that they are in the ratio 10,000 : 4000: 1000. That is, in the ratio 10:4:1. If we want a sample of size 30,20 $\left( = \frac{30}{15} \times 10 \right)$ items are drawn from low income group, 8 $\left( = \frac{30}{15} \times 4 \right)$ items are drawn from medium income group and 2 $\left( = \frac{30}{15} \times 1 \right)$ items are drawn from high income group. In this method we give importance to each stratum according to its size.

***(ii) Allocation proportional to variations Within each stratum :*** Here we consider the standard deviation of items belonging to each stratum and items are drawn in proportion to the size of the standard deviation of the stratrum. For instance, suppose 15,10 and 5 are the standard deviations of the income of people belonging to low medium and high income groups respectively. Now items are drawn from each group in such a way that they are in the ratio 15 : 10:5 (ie) in the ratio 3:2:1. So, to get a sample of size 30, number of items to be drawn from low income group is

$\frac{30}{6} \times 3 = 15$; number of items to be drawn from medium income group is $\frac{30}{6} \times 2 = 10$;

and from high income group is $\frac{30}{6} \times 1 = 5$

This procedure calls for knowledge of the standard deviation of the different strata into which the population is to be divided. This is the main difficulty with this procedure. Even if the standard deviations of different strata are known, this principle will be satisfactory by itself, the number of items belonging to various strata are equal or approximately so. But number of items belonging to various strata are seldom equal. Hence, we need a method which will take into account both the size and variations of each stratum.

***(iii) Optimum allocation :*** This method is a combination of the two principles (i) and (ii) given above. Here both the size of each stratum and the variation within each stratum are taken into account in order to decide the number of items to be drawn from each stratum. In this method items are drawn in proportion to the product of the total number of items and the standard deviation of each stratum.

After the method of representing each stratum has been decided upon, items are selected from each stratum by unrestricted random sampling method to get our required stratified sample.

Stratified sampling gives more accurate results compared to unrestricted random sampling.

## Cluster Sampling

As in the case of stratified sampling, here also the universe is divided into groups. But the method of dividing the universe into groups is just opposite to that in stratified sampling. In stratified sampling the groups are such that (i) there is as much homogeneity as possible with in each group and (ii) there is as marked a difference as possible between the groups. But in cluster sampling the universe is divided into groups or clusters such that (i) there is as much heterogeneity as possible with in each cluster and (ii) as much difference as possible between the cluster.

Suppose we want to conduct an enquiry into the wages earned by workers in an industry. The industry may consist of several companies. The industry is our universe and is divided into groups, each group being a company. Unlike in the case of stratified sampling, in cluster sampling we select few clusters or groups by unrestricted random sampling method. In our example, few companies are selected by unrestricted random sampling method. The clusters (ie. companies, in our example) are called primary sampling units.

There may be a number of employees in each company. These employees are called elementary sampling units. Now, we can either investigate each and every elementary sampling unit in each selected cluster or we can draw an unrestricted random sample of elementary sampling units from each selected cluster. That is we take into account each and every employee in each selected company or we can draw a sample of employees from each selected company by unrestricted random sampling method.

Clusters formed on a geographical basis are of great practical importance. Suppose we want to study grocery store sales in Tamil Nadu. The elementary sampling unit is grocery store. Tamil Nadu may be divided into a number of districts and few districts may be selected at random. Thus, the districts form the primary sampling units. Random sample of grocery stores from each selected district may be drawn and studied.

As a general rule, a cluster sample is considerably less expensive than an unrestricted random sample of equal size. It saves time also. In the above illustration the interviewer has to travel between the locations of elementary sampling units (ie. grocery stores). Considerable amount of money and time will be saved if cluster sampling is used rather than unrestricted random sampling method.

Further more, lack of complete and up to-date list of elementary sampling units makes it impossible to use unrestricted random sampling method., In such cases cluster sampling will be the only practical solution.

There is however, an important limitation of cluster sampling method, (like) this method generally gives less accurate results compared to other methods of sampling.

## Systematic sampling

While stratification and cluster sampling involve the division of the universe into groups, systematic sampling involves an ordering of the universe. The ordering may be alphabetical, numerical, geographical, or any other.

After the items in the universe have been ordered, every 10th or 20th or 30th items are selected depending upon the size of the universe and the size of the sample to be selected.

For example, suppose there are 150 rooms in a hostel each with a serial number and we have to select a sample of 10 students out of the 150 students residing in these rooms. Here we can select the students in every 15th room. The students may be arranged in another way also. The name of the students may be written in alphabetical order. From this arrangement also every 15th student may be selected and thus we can get the sample. The number 15 is obtained by dividing the number of items in the universe viz. 150 by the size of sample to be selected viz. 10. This number 15 is called sampling interval.

Now the next question is from where to start the sample selection. We may start sample selection anywhere between the 1st and 15th student in the arrangement. A random start is always preferable. Hence, a number is selected at random between 1 and 15.. Suppose the

number selected is 6. Now, we start selection from the 6th student. This is, the students corresponding to the numbers 6,21,36,51.... are selected to complete the sample, systematic sample will be a random sample for all practical purposes. Hence it is also known as quasi random sample.

The main advantage of systematic sampling is its simplicity. An advantage of stratification can also be realized by an appropriate ordering of stratification.

However, systematic sampling should not be used under the situation explained below:

Suppose we have to select a sample of households in a block. Also suppose that every eighth house in the block happens to be a big house on the corner. Here if the sampling interval also happens to be eight, then the stematic sampling method would yield a sample of only big corner houses. This sample will not serve our purpose.

Thus, if the items in the population at the end of each of a particular interval are alike, systematic sampling method should not be adopted.

Upto this point our discussions on sampling have referred to random or probability sampling. We now turn to the other method of sampling known as purposive sampling.

## Multi-stage sampling

Under this method, the sample is prepared in stages. The population is divided into a number of large sampling units, each of which in turn is divided into smaller units and so on. Suppose a sample of 1000 urban households from all over the state is to be selected. In such a case, first few districts say 10, are selected. In the second stage, two cities from each district is selected. Finally, 50 households are selected from each selected city. The number of stages is multistage sampling varies depending on the convenience and the availability of suitable sampling frames at different stages.

Ordinarily multi-stage sampling is applied in big cities leading to a considerable large geographical area.

## Area Sampling

Area sampling is a form of multi-stage sampling in which maps rather than list or registers are used in sampling frame. In simple random sampling, stratified sampling and systematic sampling a complete and accurate listing of the elements of population is required. If such satisfactory frames are not available area sampling is applied. In area sampling, the overall area to be covered in a survey is divided into smaller areas within which a random sample is selected. For example, a city map can be used for area sampling. Various blocks can be identified on the map and this can provide a suitable frame. The entire city area can be divided into blocks which are them numbered and from which a random sample is finally drawn. Area sampling is perhaps the only possibility, if a suitable sampling frame is not available.

## Sequential sampling

In sequential sampling, a number of samples are randomly drawn from the population., Generally, the first sample is the largest, the second is smaller than the first, the third is smaller than the second and so on., The result of the first sample is compared and evaluated. If they do not satisfy the criteria, the second sample are computed and evaluated. If they do not satisfy the criteria, the third sample are computed. This process is continued in till the specified criteria are satisfied.

The main advantage of sequential sampling is that it obviates the need for determining a fixed sample size before commencement of the survey.

## 6. Non-probability sampling

In non-probability sampling, the selection of a population èlement is based on the judgement of the researcher. The following are the non-probability sampling methods.

1) Convenience sampling
2) Judgement sampling
3) Quota sampling

## Convenience sampling

The convenience sampling, as the name implies, is selected on the basis of the convenience of the researcher. The convenience may be in respect of availability of source list, successionalitily of the units etc. A convenience sampling may be used when ;

i) The universe is not clearly defined.
ii) the sampling unit is not clear.
iii) a complete source list is not available

In convenience sampling, it is not possible to know the representativeness of the selected sample. As such it introduces an unknown degree of bias in the estimate. So we can not make any definite result from such a sample. However, convenience samples can be justified at the exploratory stages of research, as a basis for generating hypothesis and for conclusive studies where the manager of a firm is willing to accept the risk that the study result might have been great accuracies convenience sampling is extensively used in practice.

## Judgement Sampling

Judgement samples are also called purpose samples. They are selected on the basis of what some expert thinks those particular sampling units will contribute to answering the particular research at hand. The reliability of the sample depends upon the judgement of the expert. For example, in test marketing a judgement is made as to which cities would constitute the best ones for testing the market, ability of a new product.

The judgement sample may be suitable when a small sample of a few units is to be selected as the errors of judgements are likely to be less than the random errors of probability sample.

## Judgement sampling (Non - Probability sampling)

While choosing the purposive sample only the average items are considered and extreme items are omitted. In accordance with the object of the enquiry the selection of the sample is adjusted so that no significant item can be ignored.

It is clear that in this method the bias of the investigator can play a very important role and destroy the representiveness of the sample. But there may be cases where a purposive sample may give better results than a random sample. For example, if we have to select only 5 students out of 5000 with a view to study their height measurements purposive selection can give better results than random selection. The reason is that the investigator would select such five students who have normal or mean height according to his judgement. But, however, if the size of the sample is larger, random sample would give better results than purposive sample. This is due to the law of inertia of large numbers.

## Quota sampling

It is a type of purposive sample. Quota sample, is basically a stratified purposive sample. In a quota sample, the total sample is broken down on the basis of known strata in the universe (for instance age, sex, occupation, social class etc.) Then quotas (that is number of people to be interviewed from each group) are fixed according to the proportion of people belonging to different strata in the universe leaving the selection of actual respondents to the complete discretion of the interviewers. For example, in a radio listening survey, the interviewers may be asked to interview 100 persons living in a locality and that out of these 100 persons 50 must be housewives, 30 must be farmers and 20 must be children below the age of 15. With these quotas fixed, the interviewer is free to select the persons to be interviewed.

In this method, the cost per person interviewed will be very small. But considerable amount of personal bias is likely to be present. For example, the interviewer may miss farmers working in the fields or talk with those housewives who are at home. If a person refuses to respond, the interviewer may simply select some other person.

Quota sampling method is very popular in market surveys and public opinion polls.

## Sample size

Determination of sample size involves both statistical and non-statistical considerations. They are:

### 1. Nature of Population

If all the units in the population are homogeneous, a small sample is enough. If they are heterogenous, a large sample would be needed, Variability in the population can be known by estimating its standard deviation. Sometimes, estimates of standard deviaion are available from previous studies.

## 2. Degree of accuracy or precision desired.

Precision is the amount of error that could be tolerated in sample estimates. In general, the larger the sample, the smaller the error or higher the degree of accuracy. The accuracy of a sample increase in proportion to increase in the square root of the number of items in the sample. To double the accuracy, we must quadruple the size of the sample.

## 3. Degree of confidence desired

The third consideration is the degree of confidence that the researcher wants to have in the results of the study. In case, he wants to be 100 per cent confident of the results he is left with no option but to cover the entire population. Normally, three confidence levels namely, 99 percent, 95 percent and 90 percent are used. In case, 95 percent confidence level is used, it implies 5% of the true population statistically falling outside the range indicated by the confidence level. In marketing research studies, the most frequently used norm is the 5% confidence level.

## Non-statistical considerations

1) **Nature of study :** For intensive and longitudinal studies small samples are enough.

2) **Practical consideration :** Availability of finance and time at our disposal, number of trained field workers etc., are important factors in deciding the size of the sample.

3) **Nature of respondents to be contacted :** If respondents are geographically scattered, a small sample is suitable. If refusal rate is likely to be heavy, a large sample has to be selected.

No rigid number can be prescribed for an optimum size of sample. The nature of the problem is the only deciding factor and common sense and experience of the researcher are the only guide in this matter.

Cost and time are important factor in practical sampling work. Purposive sampling often requires less outlay and less total time compared to probability sampling.

In pilot studies often purposive samples are used by the statisticians.

## Sampling Distribution

Any sample taken from a population is not studied for its own sake. Instead, we are interested in sample information because of the following two reasons :

(I) **For estimation :** That is, to use the characteristics such as the mean and standard deviation of the sample to estimate the corresponding characteristics of the population.

(II) **For testing hypothesis :** That is, to use it to test some hypothesis about the population from which the sample has been drawn.

The whole sampling theory deals with only the above two problems namely, estimation and testing of hypothesis.

From the standpoing of sampling theory we are concerned with three different frequency distributions, their characteristics and relations. These three frequency distributions are :

(i) the universe distribution (ii) the sample distributions, and (iii) the sampling distribution. Now let us see these there distributions one by one.

**The Universe Distribution :** Suppose we want to study how the workers are paid in all the industries of Tamil Nadu and we have complete records on the salaries paid to all the workers in all the industries. That is, we know the values of all the items of the population. If we construct a frequency distribution of these values, it is the universe distribution. For this frequency distribution, we can calculate the mean, standard deviation etc. This mean value is called the population mean or the mean of the universe or true mean. Population mean is usually denoted by the symbol $\mu$ (read as mu). The standard deviation of the universe distribution is called the population standard deviation and is denoted by the usual symbol $\sigma$ (read as sigma). Measures characterising a universe such as $\mu$ and $\sigma$ are called parameters.

**The Sample Distribution :** In the above example suppose we do not have complete records on the salaries paid to all the workers in all the industries of Tamil Nadu. Suppose we select a random sample of few industries and collect the salary figures of the workers working in these industries. We can construct frequency distribution using these sample salary figures. This frequency distribution is called the sample distribution. For this sample distribution, we can calculate mean, standard deviation etc. This mean is denoted by the usual symbol, $\overline{x}$ and the standard deviation is denoted by the letter s. A measure characterizing a sample, such as $\overline{x}$ of s, is called a statistic.

**The Sampling Distribution :** Suppose we have a large population consisting of 5000 items. Suppose the population mean $\mu$ is 10 and the population standard deviation, is $\sigma$ 3, let us draw a random sample of some 50 items from the given population, 50 is called the size of the sample suppose the mean of sample is 10.06. That is, the value of x is 10.06. Suppose we take another sample of the same size 50. When the second sample is drawn the items of the first sample are also included in the population items so that the population always contains the same 5000 items, suppose we calculate the mean of the 3 second sample and let it be 10.02. This another value of $\overline{x}$. In the same way, suppose we take repeatedly a large number of such samples of the size 50 from the same population. The means of these samples be 10.0, 9.07, 9.09, 10.05, 10.0, 9.02 and so on. If we have taken 100 samples we will have 100 samples means. That is we will have 100 values of $\overline{x}$ . Using these values of $\overline{x}$ we can construct a frequency distribution. Such a distribution of sample means is called a sampling distribution. Here, the sample means constitute the raw material out of which the sampling distribution is constructed. Hence it is also known as the sampling distribution of mean. The sample means which are the items of the sampling distribution will cluster very closely around the mean of the universe. Hence, the mean of this sampling distribution will be the same as the mean of the universe $\mu$ . In our example 10 will be the

mean of the sampling distribution. Hence the mean of the sampling distribution may be designated by the same symbol as that of the mean of the universe involved namely $\mu$. The standard deviation of the sampling distribution is given a special name standard error. In this example as we have used the sample means to construct the sampling distribution, the standard error is also known as the standard error of mean and is denoted by the symbol $\sigma$.

Just like the sample means, we can also calculate the sample standard deviations. That is, we can calculate the value **s** for each sample. Using these values of s also, we can construct a frequency distribution. Such a distribution formed out of the sample standard deviations is called the sampling distribution of the standard deviation., The sample standard deviation will cluster closely around the standard deviation of the universe $\sigma$. Hence, the mean of sampling distribution of the standard deviation will be the same as $\sigma$. The standard deviation of this sampling distribution is called the standard error of standard deviation and is denoted by σs. In a similar manner, we can also construct sampling distributions of other statistics such as median, mean deviation etc.

Now we can define the sampling distribution as follows.

If we select a number of independent random samples of a definite size from a given population and calculate some statistics (like the mean or standard deviation) from each sample we shall get a series of values of this statistic. These values obtained from the different samples can be put in the form of a frequency distribution. The distribution so formed of all possible values of a statistic is called the sampling distribution of that statistic.

The sampling distribution of a statistic reveals certain important features:

i) A sampling distribution is generated from a universe distribution known or assumed.

ii) By varying the sample size, we can get an infinite number of sampling distributions of the same statistic for the same population. For example we can get a sampling distribution of the mean when the sample size is 50. By varying the size of the samples from 50 to 60, we can get another sampling distribution of the mean for the same population and so on.

iii) A population may generate sampling distributions or two or more different statistics. For the same population we can get sampling distribution of mean sampling distribution or median sampling distribution of standard deviation etc.

Sampling distributions are of supreme importance in theory and practice of statistics. It is because the sampling distribution of a statistic has well-defined properties and it is from these properties that we can calculate errors due to changes involved in making generalisation about populations on the basis of samples.

## Standard Error of Mean

The standard deviation of the sampling distribution of mean is known as the standard error of mean. The standard error may be considered the key to the sampling theory : If we know the standard deviation of the universe, namely **σ**, the standard error for mean can be found out with

the help of σ itself; we need not take a number of independent random samples, form the sampling distribution of mean and find out its standard deviation (i.e) the standard error of mean. It has been found that the standard error of mean is equal to the universe standard deviation divided by the square root of the sample size.

That is $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

Where $\sigma_{\bar{x}}$ is the standard error of mean.

σ is the universe standard deviation, and n is the size the sample.

For example, suppose we have selected a random sample consisting of 50 items from a population whose standard deviation is equal to 4. Here we can get the standard error of mean as follows.

| | | | |
|---|---|---|---|
| We are given | n | = | size of the sample |
| | | = | 50 |
| | σ | = | universe standard deviation |
| | | = | 4 |

∴ The standard error of mean

$$\sigma_{\bar{x}} = \frac{\sigma}{n}$$

$$= \frac{4}{50}$$

$$= .5657$$

## Properties of the standard error of mean.

1) The standard error of mean is always smaller than the universe standard deviation. It is obvious from the formula given above namely $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ . This fact can also be made clear in another way. There may be some very large and very small values in the universe in the example of salaries of industrial workers in Tamil Nadu. A few workers may get salaries as low as Rs.30 per month where as a few others may get as high as Rs.800 per month. The remaining workers may get in between these two figures. Thise wide dispersion will affect the magnitude of

the universe standard deviation whereas it will not affect the magnitude of the standard error so much. Because the standard error is only a measure of variability in a distribution of means. If we draw equal size random samples, their means will differ but no mean will be equal to Rs.800 or to Rs.30. Thus a distribution of sample means will show greater uniformity than the universe distribution. It is reflected in the formula for the standard error.

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

2) The larger the sample size, the smaller the standard error and the smaller the sample size the larger will be the standard error. For instance, if the sample size is 100.

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{100}} = \frac{\sigma}{10}$$ On the other hand, if the sample size is 400

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{400}} = \frac{\sigma}{\sqrt{20}}$$

This variation in the standard error according to the variation in sample size can also be demonstrated by the following illustrative hypothesis.

Suppose we draw samples of the same size as the universe. That is, if the universe consists of 5000 items, all the 5000 items are taken as the sample each time. In this case the mean of each sample would be identical with the true mean. In such an extreme case the 'sampling distribution' would consist of means each of which is identical with the true mean and the standard error would be zero. On the other hand, suppose a large number of samples of say, 50 items each is drawn from the population. Some of these samples may have the chance inclusion of an undue number of small or large items which are not offset. Hence, the means of these samples may not be close to the true mean. Such widening of the dispersion of sample means about the true mean is reflected in the larger standard error.

3) It follows from the formula for the standard error that larger the variation in the universe distribution, the larger the variation in the distribution of sample means and vice versa. That is larger the value of $\sigma$, the larger value of $\sigma_{\bar{x}}$ and vice versa.

So far we have assumed that the standard deviation of the universe is known. However, in practice, the universe standard deviation is rarely known. When the universe standard deviation is not known we cannot use the formula $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ to get the value of standard error. Since the very purpose of sampling is to avoid having to analyse the entire universe, for practical work, we take the standard deviation of a sample as an estimate of the value of $\sigma$. Consequently

to obtain the standard error for practical work it is enough if we know the standard deviation and the size of one single sample. Now, the formula for the standard error takes the following form.

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

If by chance, we know several sample means from the sampling distribution, the standard deviation of these several means from their mean may be taken as an estimate of the standard error.

## Uses of the standard error of the mean

1) Standard deviation is a measure of the variation of individual items about their mean values. Here, in the sampling distribution, the individual items are the sample means and the standard error measures the variation of means, from sample about the true means. The variation of sample means from the true means is called the sampling error and is due to the effect of chance in sampling. For example, if we take five random samples these five mean values will differ from one another because of chance. These chance variations or sampling variations can be measured with the help of standard error.

2) Just as the standard deviation is used to locate an item with a distribution so the standard error of the mean tells us the location of a sample mean relative to the true mean.

3) Other things being equal, it is desirable to have a standard error as small as possible. Smaller the standard error, better is the chance for a single mean to lie close to the true mean. In such a case, the single mean can be used as a good estimate of the true mean.

## Model Questions

1. Explain the nature and merits of sampling.
2. Describe the various types of sampling methods.

## LESSON - 11
# TESTING OF HYPOTHESIS

The object of the present lesson is to describe test of significance of statistical hypothesis. There can be several types of hypotheses. For example, a coin may be thrown 200 times and we may get heads 80 times and tails 120 times. We may now be interested in testing the hypothesis that the coin is unbiased. To take another example we may study the average weight of 100 students of a particular college and may get the result of 110 lbs. We may now be interested in testing the hypothesis that the sample has been drawn from a population with average weight 115 lbs. Similarly we may be interested in testing the hypothesis that the variables in the population are uncorrelated. The procedure adopted in testing a hypothesis is as follows:

### Procedure of testing a Hypothesis

### 1. Set up a Hypothesis

A hypothesis is a supposition made as a basis for reasoning. All scientific theories are tested for setting up a hypothesis against data of observation. If the facts are not inconsistent with a given hypothesis, it must be rejected. If the facts are not inconsistent with the hypothesis, the hypothesis is tenable. The test of significance may best be oriented logically around the statement of a hypothesis with which the data may or may not conform. A statistical hypothesis is a hypothesis concerning the parameters or form of the probability distribution for a particular population. Various kinds of hypotheses are possible. For example, one value is larger than another, that one value is smaller than another, or that the two values are not equal.

The conventional approach to hypothesis testing is to construct a single hypothesis about the population parameter but rather to set up two different hypotheses. These hypotheses must be constructed so that if one hypothesis is true, the other is false and vice versa.

The two hypotheses in a statistical test are normally referred to as :

1. Null hypothesis and
2. Alternative hypothesis

The null hypothesis is a very useful tool in testing the significance of difference. In its simplest form, the hypothesis asserts that there is no true difference in the sample and the population in the particular matter under consideration and that the difference found is accidental and unimportant arising out of fluctuations of sampling. The null hypothesis is akin to the legal principle that a man is innocent until he is proved guilty. It constitutes a challenge and the function of the experiment is to give the facts a chance to refute his challenge. For example, if we want to find out whether extra coaching has benefited the students or not, we shall set up a null hypothesis that "extra coaching has not benefited the students". Similarly, if we want to find out whether a particular drug is effective in curing Malaria", the rejection of null hypothesis indicates that the differences have statistical significance and the acceptance of the null hypothesis indicates that

the differences are due to chance. Since many practical problems aim at establishing statistical significance of differences, rejection of the null hypothesis may thus indicate success in statistical project.

As against the null hypothesis, the alternative hypothesis specified those values that the researcher believes to hold true, and of course he hopes that the sample data lead to acceptance of this hypothesis can be tested at one time against the null hypothesis. Now a days, it is usually accepted common practice not to associate any special meaning to the null or alternative hypothesis but merely to let these terms represent two different assumptions about the population parameter. However, for statistical convenience it will make a differences as to which hypothesis is called the null hypothesis and which is called the alternative.

The null and alternative hypothesis are distinguished by the use of two different symbols. $H_0$ representing the null hypothesis and has the alternative hypothesis. Thus a psychologist who wishes to test whether or not a certain class of people have a mean I.Q. higher than 100 might establish the following null and alternative hypotheses:

Or, if he wishes to test for difference between the mean I.Q. of two groups, this psychologist might want to establish the null hypothesis that the two groups have equal means ($\mu_1 - \mu_2 = O$) and the alternative hypothesis that their means are not equal.

($\mu_1 - \mu_2 \neq O$)

Ho : $\mu_1 - \mu_2 = O$ (null hypothesis)

Ha : $\mu_1 - \mu_2 \neq O$ (alternative hypothesis)

## 2. Set up a suitable significance level

Having set up the hypothesis the next step is to test validity of $H_0$ against that of Ha at certain level of significance. The confidence with which an experimenter rejects or retains a null hypothesis depends upon the significance level adopted. This significance level is customarily expressed as a percentage, such as 5 percent, is the probability of rejecting the null hypothesis if it is true, When the hypothesis in question is accepted at the 5 percent level, the statistician is running the risk that, in the longrun, he will be making the wrong decision about 5 percent of the time. By rejecting the hypothesis at the same level he runs the risk of rejecting a true hypothesis in 5 out of every 100 occasions. By testing at the 1 percent level he seeks to reduce the chance of making a false judgement but some element of risk remains ( 1 out of 100 occasions) that he will make the wrong decision. ie. he may accept where he ought to have rejected or vice versa.

## 3. Setting a test Criterion

The third step in general testing procedure is to construct a test criterion. This involves selecting an appropriate probability distribution for the particular test, that is, probability distribution that are commonly used in testing procedures t, F and $x^2$. Test criteria must employ an appropriate probability distribution; for example, if only small sample information is available, the use of the normal distribution would be inappropriate.

## 4. Doing computations

Having taken the first three steps, we have completely designed a statistical test. We now proceed to the fourth step performance of various computations from a random sample of size n, necessary for the best. These calculations include the testing and the standard error of testing statistic.

## 5. Making Decision

Finally, as a fifth step, we may draw statistical conclusions and may make decisions. A statistical conclusion or statistical decision is a conclusion either to reject or not to reject the null hypothesis. The decision will depend on whether the computed value of the test criterion falls in the region of rejection or the region of acceptance. If the hypothesis is being tested at 5% level and the observed set of result has probabilities less than 5 percent, we consider the difference between the sample statistics and the hypothetical parameter significant. In other words, we think that the sample result is so rare that it cannot be explained by chance variation alone. We then decide to reject Ho and state : " the null hypothesis is false", or the sample observations are not consistent with the null hypothesis (the rejection of Ho automatically leads to acceptance of Ha)

On the other hand, if at 5% level of significance, the observed set of results has probability more than 5% we give reason that the difference between the sample result and the hypothetical parameter can be explained by chance variations and therefore is not significant statistically. Consequently, we decide not to reject Ho and state : " The sample result is not inconsistent with the hypothesis" if the probability is about 5 percent, the wisest course may be to resolve judgement and draw another sample if possible.

## Two types of Errors in Testing of Hypothesis

When a statistical hypothesis is tested there are four possible results:

1. The hypothesis is true but our test rejects it.
2. The hypothesis is false but our test accepts it.
3. The hypothesis is true and our test accepts it.
4. The hypothesis is false and our test rejects it.

Obviously, the first two possibilities lead to errors. When the result of the test leads to the rejection of a null hypothesis which is true (possibility No 1) this is called type I error.

In other words, type I errors are made when we reject a null hypothesis by making a difference significant, although no true difference exists. On the other hand, if the result of a statistical test leads to the acceptance of the null hypothesis which is not true this is called type II error. In other words, type II errors are made when we accept a null hypothesis by making a difference not significant when a true difference actually exists. The distinction between these two kinds of errors can be made clear by an example. Assume that the difference between two population means is actually zero. In our test of significance when applied to the two sample means leads us

to believe that the difference in population means is significant, we make a type I error. On the other hand, suppose there is a true difference between the two population means. Now if our test of significance leads to the judgement "not significant" we commit a type II error. We thus find ourselves in the situation which is described by the following table:

| | Accept H | Reject H |
|---|---|---|
| | Correct decision | Type I error |
| H is false | Type II Error | Correct decision |

The probability of making type I error is fixed in advance by the choice of level of significance employed in the test. At the 5 percent level we know that we shall be rejecting on the average 5 percent of all true hypothesis, because even if the hypothesis were true, [T], for example, would be greater than, 1.96 just due to chance 5 times out of 100. It may be noted that we can make the type I error as small as we please by lowering level of significance. For example, if we work with, 1 percent level of significance the critical value of [T] raise from 1.96 to 2.58. But by so doing we increase the chance of accepting a false hypothesis ie., of making a type II error. It follows that it is impossible to minimise both errors simultaneously.

Hence it appears that an experimenter must decide which kind of wrong inference he would rather avoid as apparently he can prevent one type of error only at the risk of making the error more likely. In the long run, errors of type I are perhaps more likely to prove serious in research programme in the Social Sciences than are errors of type II. However, an appropriate balance to be given to the two types of error in designing tests of significance will depend upon the practical consequences which follow from making errors of both kinds. As a general rule it is probably wise to demand a significance level 3.01 in most experimental research, ie. to risk type II errors by preventing those of type I. But the 0.05 level is often satisfactory, especially in preliminary work.

Having discussed the above concepts let us now discuss the various situations where we have to apply the various tests of significance. For the sake of convenience and clarity these situations may be summed up under the following three heads:

I. Sampling of Attributes

II. Sampling of variables (Large samples)

III. Sampling of variables ( Small Samples)

## I. Sampling of Attributes

As distinguished from variables where quantitative measurement of a phenomenon is possible, in case of attributes we can find out the presence or absence of a particular characteristic. The sampling of attributes may, therefore, be regarded as the drawing of samples from a population whose members possess the attribute A or not A. For example, in the study of attribute "Literacy" a sample may be taken and people classified as literates and illiterates. With such data the binomial type of problem may be formed. The selection of an individual on sampling may be called "event" the appearance of an attribute A may be taken as 'success' and its non appearance as 'failure'. Thus if out of 1000 people selected for the sample, 100 are found literates and 900 illiterates, we would say that the sample consists of 1000 events out of which 100 are successes and 900 failures. The probability of success of p $\frac{100}{1000}$ or 0.1 and the probability of failure or $q = \frac{900}{1000}$ 0.9. So that p + q = 0.1 + 0.9 = 1.

The sampling distribution of number of successes being a binomial probability model, would have as its mean $\mu = np$ and as its variance $\sigma^2 = npq$ or standard Deviation $\sigma = \sqrt{npq}$

### Standard error of Number of successes

If we have N sample with n events each, the chance of success in each event is p and of its failure 1, ie. (I -p). The standard error of the success is given by the formula.

$$\text{SE of number successes} = \sqrt{npq}$$

### Standard Error of the proportion of success

Instead of recording the number of successes in each sample we might record the proportion of successes, that is, $\frac{1}{n}$ of the number in each sample. As this would amount to dividing all figures for the record by n, the mean proportion of successes must be, p and the standard deviation of the proportion of successes are

$\frac{\sqrt{pq}}{n}$ thus we have the following formula.

$$S\ Ep = \frac{\sqrt{pq}}{n}$$

### Illustration : 1

Tax authorities feel that 15% of the families in a city have an annual income of Rs.10000 or more. A random sample of 100 families in the city yields the fact that only 7% have incomes of Rs. 10,000 or more. Using 5% level of significance, test the hypothesis that p = 15%

## Solution

Let us take the hypothesis that the proportion of families having an annual income of Rs.10000 or more p = 15% = 0.15.

q = 85% = .85

$$\text{Standard Error} = \frac{\sqrt{pq}}{n} = \frac{(.15)\,(.85)}{100} = .036$$

difference between population proportion and sample proportion = 15 - .07

= .08

$$\frac{\text{Difference}}{\text{S.E.}} = \frac{08}{.036} = 2.22$$

Since the difference is more than 1.96 SE (5% level) there is a significant difference between population proportion and sample proportion.

## Standard Error of the Difference between the proportions

If two samples are drawn from different populations, we may be interested in finding out whether the difference between the proportion of successes is significant or not. In such a case we take the hypothesis that difference between $p_1$ ie. the proportion of successes on one sample and $p_2$ ie. the proportion of successes in another sample is due to fluctuations of random sampling. The standard error of the difference between proportions is calculated by applying the following formula :

$$\text{S.E. } (p_1 - p_2) = \sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

Where p = the best estimate of the actual proportion in the population. The value of p is obtained as follows :

$$p = \frac{n_1p_1 + n_2p_2}{n_1 + n_2}$$

If $\frac{p_1 - p_2}{SE}$ is less than 1.95 SE (5% level ) the difference is regarded as due to random sampling variation, ie. as not significant.

The following illustration shall explain the method.

### Illustration -2

In a random sample of 1000 persons from town A 400 are found to be consumers of wheat. In a sample of 800 from town B, 400 are found to be consumers of wheat. Do these data reveal a significant difference between town A and town B, so far as the proportion of wheat consumers is concerned?

## Solution

Let us take the hypothesis that the two towns do not differ so far as proportion of wheat consumers is concerned.

Computing the standard error of the difference of proportions

S.E.

$$(p_1 \ p_2) = \sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

$$n_1 = 1000, \qquad p_1 = \frac{400}{1000} = 0.4$$

$$n_2 = 800 \qquad p_2 = \frac{400}{800} = 0.5$$

$$p = \frac{(1000 \text{ x } 0.4) + (800 \text{ x } 0.5)}{100 + 800} = \frac{4}{9}$$

$$q = 1 - \frac{4}{9} = \frac{5}{9}$$

$$\text{S.E. } (p_1 - p_2) = \sqrt{\frac{4}{9} \text{ x } \frac{5}{9} \quad \frac{1}{1000} + \frac{1}{800}}$$

$$= \sqrt{\frac{20}{81} \text{ x } \frac{9}{4000}}$$

$$= 0.024$$

$(p_1 - p_2) = 0.4 - 0.5 = -0.1$*

$P_1 - P_2 = 0.4 - 0.5 = -0.1$. However the conclusion remains the same irrespective of the fact whether the difference is positive or negative.

$$\frac{\text{Difference}}{\text{S.E.}} = \frac{0.1}{0.024} = 4.17$$

## Large samples and small samples

It is difficult to draw a clear -cut line of demarcation between large and small samples, it is normally agreed amongst statisticians that the sample is to be as large only if its size exceeds 30. The tests of significance used for dealing with problems relating to large samples are different from the one used for small samples for the reason that the assumptions that we make in case of large samples do not hold good for small samples. The assumptions made while dealing with problems relating to large samples are :

(i) The random sampling distribution of a statistic is approximately normal ; and

(ii) Values given by the samples are sufficiently close to the population value and can be used in its place for the standard error of the estimate.

### 1) Standard error or Mean

(i) When standard deviation of the population is known

$$S.E. \ \overline{x} = \frac{\sigma}{\sqrt{n}}$$

Where SE $\overline{x}$ refers to standard error of mean

$\sigma$ = standard deviation of the population

n = Number of observations in the sample

(ii) When the standard deviation of the population is not known we have to use standard deviation of the sample in calculating standard error of mean. Consequently the formula for calculating standard error is

$$S.E. \ \overline{x} = \frac{\sigma \text{ (sample)}}{\sqrt{n}}$$

Where $\sigma$ denotes standard deviation of the sample

It should be noted that if standard deviation of both sample as well as population are available then we must prefer standard deviation of the population for calculating standard error of mean.

### 2) Standard Error of Median $= 1.2533 \times \frac{\sigma}{\sqrt{n}}$

Standard Error of a standard Deviation or

$$S.E. \ \sigma = \frac{\sigma}{\sqrt{2n}}$$

3) Standard Error of Coefficient of Correlation or

$$S.E._r = \frac{1 - r^2}{\sqrt{n}}$$

## Illustration : 3

A sample of 100 students is taken from a large population. The mean height of these students is 64 inches and the standard deviation 4 inches. Can it reasonably be regarded that in the population mean height is 66 inches?

## Solution

We have to test the hypothesis that the mean height in the population is 66 inches.

$$\text{S.E. of mean} = \frac{\sigma}{\sqrt{n}} = \frac{4}{\sqrt{100}} = \frac{4}{\sqrt{10}} = 0.4$$

Difference between the sample and population mean = 66 - 64 = 2.

$$\frac{\text{Difference}}{\text{S.E.}} = \frac{2}{0.4} = 5$$

Since the difference between sample mean and population mean is more than 2.58 S.E ( 1% level of significance) the result of experiment does not support the hypothesis that the mean height in the population is 66 inches.

## Illustration

Ten individuals are chosen at random from a population and their heights are found to be (in inches) : 63,63,66,67,68,69,70,70,71 and 71. In the light of the data discuss the suggestion that the mean height in the population is 66 inches.

## Solution

We have to test the hypothesis that the mean height in the population is 66 inches. Applying t-test

$$t = \frac{(x - \mu)}{S} x\sqrt{n}$$

## CALCULATION OF MEAN ($\bar{X}$) AND STANDARD DEVIATION

| Height in inches X | X -67 d | $d^2$ |
|---|---|---|
| 63 | -4 | 16 |
| 63 | -4 | 16 |
| 66 | -1 | 1 |
| 67 | 0 | 0 |
| 67 | 0 | 0 |
| 68 | +1 | 1 |
| 69 | +2 | 4 |
| 70 | +3 | 9 |
| 71 | +4 | 16 |
| 71 | +4 | 16 |
| $\Sigma X = 678$ | $\Sigma d = 8$ | $\Sigma d^2 = 88$ |

$$\bar{x} = \frac{\Sigma X}{n} = \frac{678}{10} = 67.8$$

$$\bar{d} = \frac{\Sigma d}{n} = \frac{8}{10} = 0.8$$

$$S = \sqrt{\frac{\Sigma d^2 - (d^2) X n}{n - 1}}$$

$$= \sqrt{\frac{88 - 6.4}{9}} = \sqrt{\frac{81.6}{9}} = \sqrt{9.07} = 3.011$$

$x = 67.8$ $S = 3.011$ $\mu = 66,$ $n = 10$

Substituting the Values,

$$t = \frac{67.8 - 66}{3.011} \times \sqrt{10}$$

$$= \frac{1.8 \times 3.162}{3.011} = 1.89$$

Degrees of freedom to $\gamma = (10 - 1) = 9$

For 9 degrees of freedom the table value of t at 5% level of significant (0.05) = 2.26.

Since the calculated value of t is less than the table value, the experiment provides no reason for doubting the hypothesis that the mean height in the population is 66 inches.

## 2. To test the difference between the means of two samples

Given two independent random samples of size $n_1$ and $n_2$ with means $\overline{x}_1$ and $\overline{x}_2$ standard deviation S1 and S2. We may be interested in testing the hypothesis that the samples come from the same normal population. To carry out the test, we calculate the statistic t,

$$t = \frac{x_1 - x_2}{S} \times \sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$

Where

$x_1$ = mean of the first sample

$x_2$ = mean of the second sample

$n_1$ = number of observations in the first sample

$n_2$ = number of observations in the second sample

S = combined standard deviation

The value of S is calculated as follows :

$$S = \sqrt{\frac{\Sigma (X_1 - \overline{x}_1)^2 + \Sigma (X_2 - \overline{x}_2)^2}{n_1 + n_2 - 2}}$$

When the actual means are in fraction the deviations should be taken from assumed means. In such a case the combined standard deviation is obtained by applying the following formula.

$$S = \sqrt{\frac{\Sigma (X_1 - A_1)^2 + \Sigma (X_2 - A_2)^2 - n_1 (\overline{x}_1 - A_1)^2 - n_2 (\overline{x}_2 - A_2)^2}{n_1 + n_2 - 2}}$$

$A_1$ = Assumed mean of the first sample

$A_2$ = Assumed mean of the second sample

$x_1$ = Actual mean of the first sample

$x_2$ = Actual mean of the second sample

t here is based on $(n_1 + n_2 - 2)$ degrees of freedom.

In the calculated value of t be > to 0.5 (to .01), the difference between the sample means is said to be significant at 5% (1%) level of significance otherwise the data are said to be consistent with the hypothesis.

The following example will illustrate the test.

## Illustration - 4

For a random sample of 10 pigs, fed on diet A, the increases in weight in pounds in a certain period were :

10,6,16,17,13,12,8,14,15,9

For another random sample of 12 pigs, fed on diet B the increases in the same period were :

7,13,22,15,12,14,18,8,21,23,10,17

Test whether the diets A and B differ significantly as regards their effect on increase in weight.

## Given the following :

| Degrees of freedom | Values of "t" at 5% level |
|---|---|
| 19 | 2.09 |
| 20 | 2.09 |
| 21 | 2.08 |
| 22 | 2.07 |
| 23 | 2.07 |

## Solution

Let us take the null hypothesis that diets A and B do not differ significantly as regards their effect on increase in weight. Applying t - test

$$t = \frac{\bar{x}_1 - \bar{x}_2}{S} \times \sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$

$$S = \sqrt{\frac{\Sigma (x_1 - \bar{x}_1)^2 + (X_2 - \bar{x}_2)^2}{n_1 + n_2 - 2}}$$

**Calculating the required values :**

| Pigs fed on diet A | | | Pigs fed on diet B | | |
|---|---|---|---|---|---|
| Increase in weight $X_1$ | Deviations from mean 12 $(X_1 - \overline{x}_1)$ | $(X_1 - \overline{x}_1)^2$ | Increase in weight $X_2$ | Deviations from mean 12 $(X_2 - \overline{x}_2)$ | $(X_2 - \overline{x}_2)^2$ |
| 10 | -2 | 4 | 7 | -8 | 64 |
| 6 | -6 | 36 | 13 | -2 | 4 |
| 16 | +4 | 16 | 22 | +7 | 49 |
| 17 | +5 | 25 | 15 | 0 | 0 |
| 13 | +1 | 1 | 12 | -3 | 9 |
| 12 | 0 | 0 | 14 | -1 | 1 |
| 8 | -4 | 16 | 18 | +3 | 9 |
| 14 | +2 | 4 | 8 | -7 | 49 |
| 15 | +3 | 9 | 21 | +6 | 36 |
| 9 | -3 | 9 | 23 | +8 | 64 |
| | | | 10 | -5 | 25 |
| | | | 17 | +2 | 4 |
| $\Sigma X_1 = 120$ | $\Sigma (X_1 - \overline{x}_1) = 0$ | $\Sigma (X_1 - \overline{x}_1)^2 = 120$ | $\Sigma X_2 = 180$ | $\Sigma (X_2 - \overline{x}_2) = 0$ | $\Sigma (X_2 - \overline{x}_2)^2 = 314$ |

Mean increase in weight of 10 pigs fed on diet A

$$\overline{x}_1 = \frac{\Sigma X_1}{n_1} = \frac{120}{10} = 12 \text{ pounds}$$

Mean increase in weight of 12 pigs fed on diet B

$$\overline{x}_2 = \frac{\Sigma X_2}{n_1} = \frac{180}{12} = 15 \text{ pounds}$$

$$S = \sqrt{\frac{\Sigma (x_1 - x_1)^2 + (X_2 - x_2)^2}{n_1 + n_2 - 2}}$$

$$= \sqrt{\frac{120 + 314}{10 + 12 - 2}} = \sqrt{\frac{431}{20}} = 4.65$$

$\overline{x}_1 = 12, \overline{x}_2 = 15, n_1 = 10, n_2 = 12, S = 4.65$. Substituting the values in the above formula :

$$t = \frac{12 - 15}{4.65} \times \sqrt{\frac{10 \times 12}{10 + 12}}$$

$$= \frac{3}{4.65} \times 2.33 = 1.5$$

$\gamma = n_1 + n_2 - 2 = 10 + 12 - 2$, ie. 20

For r=20 at 5% level of significance the table value of t is 2.09. The calculated value is less than the table value and hence the experiment provides no evidence against the hypothesis. We, therefore, conclude that diets A and B do not differ significantly as regards their effect on increase in weight.

## To test the significance of an observed Correlation Co-efficient

Given a random sample from a bivariate normal population we are to test the hypothesis that the correlation coefficient of population is zero. The variables, in the population are uncorrelated, we have to apply the following test :

$$t = \frac{\gamma}{\sqrt{1 - \gamma^2}} \times \sqrt{n - 2}$$

here t is based on (n - 2) degree of freedom. If the calculated value of t exceeds to .05 for If we say that the value of $\gamma$ is significant at 5% level. If t<to.05 the data are consistent hypothesis of an uncorrelated population. The following example will illustrate the point.

### Illustration - 5

A random sample of 27 pairs of observations from a normal population gives Correlation Coefficient of 0.42. Is it likely that the variables in the population are uncorrelated.

### Solution

Let us take the hypothesis that the variables in the population are uncorrelated. Applying t test

$$t = \frac{\gamma}{\sqrt{1 - \gamma^2}} \times \sqrt{n - 2}$$

Here = 0.42, n = 27. Substituting the values

$$t = \frac{0.42}{\sqrt{1 - (0.42)2}} \times \sqrt{27 - 2}$$

$$= \frac{0.42 \times 5}{0.908} = 2.31$$

$\gamma = (n - 2) = (27 - 2) = 25$

For $\gamma = 25$, to .05 = 2.06.

The calculated value of t is greater than the table value and hence the result of the experiment does not support the hypothesis that the variables in the population are correlated.

## CHI - SQUARE TEST

The Chi-Square test is one of the simplest and most widely used non-parametric tests in statistical work. The $X^2$ distribution has very many applications in that it involves the testing of hypothesis concerning discrete or qualitative data. The Greek letter $X^2$ was first used by Karl Pearson in the year 1900. The quantity $X^2$ describes the magnitude of discrepancy between theory and observation ie., with the help of $X^2$ test we are in a position to know whether a given discrepancy between theory and observation may be attributed to chance or whether it results from the inadequacy of the theory to fit the observed facts. If $X^2$ is zero it means that the observed and expected frequencies completely coincide. The greater the discrepancy between observed and expected frequencies, the greater is the value of $X^2$. The quantity of $X^2$ is defined as :

$$x2 = \Sigma \frac{(O - E)^2}{E}$$

Where O refers to the observed frequencies and E refers t the expected frequencies.

Steps : To determine the value of $X^2$ , the step required are :

(i) Calculate the expected frequencies.

(ii) Take the difference between observed and expected frequencies and obtain the squares of these differences, ie. obtain the values of $(O - E)^2$.

(iii) Divide the quantity $(O - E)^2$ obtained in step (ii) by the expected frequency and obtain the total, $\Sigma \frac{(O - E)^2}{E}$ This gives the values of $X^2$.

The calculated value $X^2$ is compared with the table value of $X^2$ for given degrees of freedom at a certain specified level of significance. If at the stated level (generally 5% level is selected) the calculated value of $X^2$ is more than the table value of $X^2$ the difference between theory and observation is considered to be significant, i.e. it could not have arisen due to fluctuations of simple sampling. If, on the other hand, the calculated value of $X^2$ is less than the table value, the difference between theory and observation is not considered as significant, ie. it is regarded as due to fluctuations of simple sampling and hence ignored.

It should be noted that the value of $X^2$ is always positive and its upper limit is infinity. Also since $X^2$ is derived from observations, it is a statistic and not a parameter. The $X^2$ test is therefore termed non-parametric. It is one of the great advantage of this test that it involves no assumption about the form of the original distributions from which the observations come.

## Degrees of Freedom

By degrees of freedom we mean the number of classes to which the values can be assigned arbitrarily or at will without violating the restrictions or limitations placed. For example, if we are to choose any five numbers whose total is 100, we can exercise our independent choice for any four numbers only, the fifth number is fixed by virtue of the total being 100 as it must be equal to 100 minus the total of the four numbers selected. For example, if the four numbers are 20,35,15,10 the fifth number must be {100-(20+35+15+10)} Thus, though we were to choose any five numbers we could choose any four only. Our choice is reduced by one because of one condition placed in the data, ie. that of total being 100 thus there was only one restraint on our freedom our degrees of freedom were only four. If more restrictions are placed, our freedom to choose will be still curtailed. For example, if there are 10 classes and we want our frequencies to be distributed in such a manner that the number of cases, the mean and standard deviation agree with the original distribution, we have three constraints and so three degrees of freedom are lost; Hence in this case the degrees of freedom will be (10 - 3) = 7 . Thus, number of degrees of freedom is obtained by substracting from the number of classes, the number of degrees of freedom lost in fitting. Symbolically, the degrees of freedom are denoted by the symbol $\gamma$ (pronounced nu) or by df and are obtained as follows : $\gamma = n = k$ where k refers to the number of independent constraints. We have a constraint or restriction whenever observed or theoretical frequencies are made to agree with one another in some one respect in the operations that lead to the calculation of $X^2$. Thus a constraint is imposed by the condition $\Sigma$ fo = $\Sigma$ fc. In general when we fit the binomial distribution, the number of degrees of freedom is one less than the number of classes, when we fit a position distribution, the degrees of freedom are 2 less than the number of classes and we fit a normal curve, the number degrees of freedom is small by 3 than the number of classes.

In a contingency table the degrees of freedom are calculated in a slightly different manner. The marginal total of frequencies place the limit on our choice of selecting cell frequencies. The cell frequencies of all columns but one (c - 1) and of all rows but one ($\gamma$ - 1) can be assigned arbitrarily and so the number of degrees of freedom for all the cell-frequencies (c - 1) ($\gamma$ - 1) where c refers to column and refers to rows thus in a 2 X 2 table the degrees of freedom = (2 -1) (2-1) = 1. Having filled up one cell in such a table the rest of the frequencies automatically follow there is no choice for them. Similarly in a 3 x3 contingency table, the number of degrees of freedom is ( 3- 1) (3 -1)=4 and so on. It means only 4 expected frequencies need to be computed. The others are obtained by subtraction from a normal totals.

## Conditions for applying $X^2$ Test

The following conditions should be satisfied before applying the $X^2$ test :

1. In the first place, N must be reasonably large. When N is small the probability given by the $X^2$ test is too small, with the result that the $X^2$ test might lead to a hypothesis being discredited where as the exact procedure might cause one not to discredit a hypothesis. It is difficult to say exactly what constitutes largeness but as an arbitrary figure we may say that N should be at least 50, however few are the cells.

2. No theoretical cell frequency should be small. Here again it is hard to say what constitutes smallness but 5 should be regarded as the very minimum and 10 is better. In practice data is in frequently contain cell frequencies below these limits. As a rule, the difficulty may be met by amalgamating such cells into a single cell entitled '10 and over'.

3. The constraints must be linear.

## Uses of $X^2$ test

$X^2$ test is one of the simplest and the most general test known. It is applicable to a very large number of problems in practice which can be summed up under the following heads.

## 1. $X^2$ test as a test of independence

With the help of $X^2$ test we can find out whether or not two attributes are associated. Suppose we have N observations classified according to criteria. We may ask whether the criteria are relative or independent. Thus, we can find out whether quinine is effective in controlling fever or not, whether there is any association between marriage and failure, or eye colour of husband and wife, in order to test whether or not the attributes are associated we take the null hypothesis that there is no association in the attributes under study or, in other words, the two attributes are independent. If the calculated value of $X^2$ is less than the table value at a certain level of the significance (generally 5% level) we say that the result of the experiment provides no evidence for doubting the hypothesis or, in other words, the hypothesis that the attributes are not associated holds good. On the other hand, if the calculated value of $X^2$ is greater than the table value at a certain level of significance, we say that the results of the experiment do not support the hypothesis or in other words, the attributes are associated. It should be noted that $X^2$ is not a measure of the degree or form of relationship, it only tells us whether two principles of classification are or are not significantly related, without reference to any assumptions concerning the form of relationship.

The following example will illustrate the application of $X^2$ test.

### Illustration - 6

Based on information on 500 randomly selected fields about the tenancy status of the cultivators of these fields and use of fertilizers collected in an agro-economic enquiry, the following classification was noted.

| | Owned | Rented |
|---|---|---|
| Using fertilizer | 208 | 92 |
| Not Using fertilizer | 32 | 168 |

Would you conclude that owner cultivators are more inclined towards the use of fertilizers?

## Solution

Let us take the hypothesis that there is no association between the tenancy status of the cultivator and the use or fertilizers. On the basis of this hypothesis the expectation for fields owned and using fertilizers is

$$\text{Expectation of (AB)} = \frac{(A) \times (B)}{N}$$

When A denotes fields using fertilizer.

$$\text{Expectation of (AB)} = \frac{240 \times 300}{500} = 144.$$

The table of expected frequencies is :

| | | |
|---|---|---|
| 144 | 156 | 300 |
| 96 | 104 | 200 |
| 240 | 260 | 500 |

Applying the $X^2$ test

| O | E | (O - E) | $\frac{(O-E)^2}{E}$ |
|---|---|---|---|
| 208 | 144 | 4,096 | 28.41 |
| 32 | 96 | 4,096 | 42.67 |
| 92 | 156 | 4,096 | 26.26 |
| 168 | 104 | 4,096 | 39.40 |

$$\Sigma \frac{(O-E)^2}{E} = 136.77$$

$$x^2 = \Sigma \frac{(O-E)^2}{E} = 136.77$$

$\gamma = (r - 1)(c - 1) = (2 - 1)(2 - 1) = 1$

For $\gamma = 1$ $x^2$ 0.05 = 3.48

For calculated value of $X^2$ is much higher than the table value and hence the result of the experiment does not support the hypothesis. We, therefore, conclude that the owner cultivators are more inclined towards the use of fertilizers.

## 2. $X^2$ Test as a test of goodness of fit

$X^2$ test is very popularly known as the test of goodness of fit for the reason that it enables us to ascertain how well the theoretical distributions such as binomial, poison, normal etc. fit empirical distributions, ie. those obtained from sample data. When an ideal frequency curve whether normal or some other type is fitted to the data, we are interested in finding out how well this curve fits with the observed facts. A test of the concordance of the two can be made just by inspection, but such a test is obviously inadequate. Precision can be secured by applying the $x^2$ test. If the calculated value of $X^2$ is less than the table value at a certain level of significance (generally 5% level), the fit is considered to be good, ie. the divergence between the actual and expected frequencies is attributed to fluctuations of simple sampling. On the other hand, if the calculated value of $X^2$ is greater than the table value, the fit is considered to be poor, ie. it cannot be attributed to fluctuation of simple sampling rather it is due to the inadequacy of the theory to fit the observed facts.

It should be borne in mind that in repeated sampling too good a fit is just as likely as too bad a fit. When the computed chi-square value is too close to zero, we should suspect the possibility that the two sets of frequencies have been manipulated in order to source them to agree and therefore, the design of our experiment should be thoroughly checked.

The following example will illustrate the application of $X^2$ test.

## Illustration -7

Three hundred digits were chosen at random from a set of tables. The frequencies of the digits were as follows:

| Digit | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| Frequency | 28 | 29 | 33 | 31 | 21 | 35 | 32 | 30 | 31 | 25 |

Using the $X^2$ test assess the hypothesis that the digits were distributed in equal number in the table. (The 5% value of $X^2$ for 9 degree of freedom is 15.92)

## Solution

We have to test the hypothesis that the digits were distributed in equal numbers. On the basis of this hypothesis

we should expect $\frac{300}{10} = 30$ as the frequency for 0,1,2,..... digits etc.

Applying the $x^2$ test.

| O | E | (O - E) | $\frac{(O-E)^2}{E}$ |
|---|---|---|---|
| 28 | 30 | 4 | 0.133 |
| 29 | 30 | 1 | 0.033 |
| 33 | 30 | 9 | 0.300 |
| 31 | 30 | 1 | 0.033 |
| 26 | 30 | 16 | 0.533 |
| 35 | 30 | 25 | 0.833 |
| 32 | 30 | 4 | 0.133 |
| 30 | 30 | 0 | 0.000 |
| 31 | 30 | 1 | 0.033 |
| 25 | 30 | 25 | 0.833 |
| Total | 300 | 300 | $\Sigma \frac{(O-E)^2}{E} = 2.864$ |

$$x^2 = \Sigma \frac{(O-E)^2}{E} = 2.864$$

Degree of freedom = 10 - 1 = 9 = $\gamma$

For = $\gamma$ = 9, x2 = 0.05 = 16.92

The calculated value of $x^2$ is less than the table value and hence our hypothesis that the digits were distributed in equal numbers holds good.

## 3. $X^2$ test as a test of Homogeneity

The $x^2$ test of homogeneity is an extension of the chi-square test of independence. In both cases we are concerned with cross - classified data. The same testing statistic used for test of independence is used for tests of homogeneity. These two types of tests are however different in a number of ways. Firstly, they are associated with different kinds of problems. Tests of independence are concerned with the problem of whether one attribute is independent of another, while tests of homogeneity are concerned with weather different samples come from the same population. Secondly, the formula involves two or more independent samples, one from each of the impossible populations in question.

The following example will illustrate the application of $x^2$ test.

## Illustration - 8

A movie producer is bringing out a new movie. In order to map out his advertising campaign he wants to determine whether the view will appeal most to a particular age group or whether it will appeal equally to all age groups. The producer takes a random sample from persons attending a preview of the new movie, and obtain the following results :

**Age groups**

| | Under 20 | 20-39 | 40-59 | 60 and over | Total |
|---|---|---|---|---|---|
| Like the movie | 146 (a) | 28(d) | 48(g) | 28 (j) | 300 |
| Disliked the movie | 54(b) | 22(e) | 42(h) | 22(k) | 140 |
| Indifferent | 20(c) | 10(f) | 10(i) | 20(l) | 60 |
| Total | 220 | 110 | 110 | 70 | 500 |

What inference will you draw from this data?

## Solution

Let us take the hypothesis that the movie appeals to all age groups.

$$\text{Expectation of (a)} = \frac{300}{500} \times 220 = 132$$

$$\text{Expectation of (b)} = \frac{140}{500} \times 220 = 61.6$$

$$\text{Expectation of (d)} = \frac{300}{500} \times 110 = 66$$

$$\text{Expectation of (e)} = \frac{140}{500} \times 110 = 30.8$$

$$\text{Expectation of (g)} = \frac{300}{500} \times 100 = 60$$

$$\text{Expectation of (h)} = \frac{140}{500} \times 100 = 28$$

The expected frequency can be tabulated as below :

| | | | | |
|---|---|---|---|---|
| 132 | 66 | 60 | 42 | 300 |
| 6.16 | 30.8 | 28 | 19.6 | 140 |
| 26.4 | 13.2 | 12. | 8.4 | 60 |
| 220 | 110 | 100 | 70 | 500 |

**Applying $X^2$ test**

| O | E | (O - E) | $\frac{(O-E)^2}{E}$ |
|---|---|---|---|
| 146 | 13.2 | 196.00 | 1.48 |
| 54 | 61.6 | 57.76 | 0.94 |
| 20 | 26.4 | 40.96 | 1.55 |
| 28 | 66.0 | 144.00 | 2.18 |
| 22 | 30.8 | 77.44 | 2.50 |
| 10 | 13.2 | 10.24 | 0.78 |
| 48 | 60.0 | 144.00 | 2.40 |
| 42 | 28 | 196.00 | 7.00 |
| 10 | 12.0 | 4.00 | 0.00 |
| 28 | 42.0 | 196.00 | 4.67 |
| 22 | 19.6 | 5.76 | 0.29 |
| 20 | 8.4 | 134.56 | 16.02 |
| | | | $\Sigma \frac{(O-E)^2}{E} = 40.14$ |

$$x^2 = \Sigma \frac{(O-E)^2}{E} = 40.14$$

$\gamma = (r - 1)(c - 1) = (3 - 1)(4 - 1) = 6$

For $\gamma = 6$, $x^2\ 0.05 = 12.59$

The calculated value of $X^2$ is much greater than the table value. The hypothesis does not hold good and we, therefore, conclude that the movie does not appeal equally to all age groups.

## Misuse of Chi- Square Test

Probably one of the most frequently used statistics is the chi-square. Unfortunately, it is also one of the most frequency misused. It is easy to learn to compete, but its correct application is not so easily learned.

The most common mistake in the application of the chi-square statistic and yet the most critical for its correct application is the violation of independence is not to be confused with the chi-square as a test of independence. The assumption of independence refers to the individual observations or frequencies and means that the occurrence of one event has no effect upon the occurrence of any other event. Another way of stating his meaning of independence is that the probability of each events occurrence is independent of the probability of occurrence of all other event. In statistical terms we say that the joint probability of two random events is equal to the product of the probabilities of these event.

The chi-square as a test of independence refers to the statistical test of the possibility of a relationship between two variables. This is often called a test of association, the question tested is whether the frequencies observed of one category are contingent upon another category. For example, is the number of 'yes' answers to some question contingent upon the age of the respondent?

Some other sources of errors in the application of $X^2$ as revealed by a survey of all papers published in the Journal of Experimental Psychology during the years 1944 - 1946 are:

(i) Small theoretical frequencies.

(ii) Neglect frequencies of non-occurrence.

(iii) Failure to equalise the sum of the observed frequencies and the sum of the theoretical frequencies.

(iv) Intermediate theoretical frequencies.

(v) Incorrect or questionable frequencies.

(vi) use of non-frequency data

(vii) Incorrect determination of the number of degrees of freedom

(viii) Incorrect computations.

The number of applications of the chi-square test does not seem to be increasing but the number of misuses $X^2$ has become surprisingly large. You cannot simply use a statistic because you know how to calculate it; you must understand the rational behind its development and the limitations on its application imposed by the assumptions underlying it.

## Limitation on the Use of $X^2$ Test

$X^2$ test is widely used in practice. However, in order to avoid the misapplication of the test the following limitations should be kept in mind.

1. Frequencies of non-occurence should be omitted for binomial or multinomial events. For example, if 5 drugs were tried out on 5 separate groups of 200 patients each, the number of cures per drug might be shown in one-way table as follows:

**DATA FOR DRUGS**

| | Drug | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | Total |
| Number Cured | 80 | 120 | 40 | 60 | 20 | 320 |

However, the $X^2$ test should not be applied to these data until the alternative outcome (ie. 'not cured') is represented in the table.

2. The formula presented here is not appropriate for cases in which repeated measurements on the same or matched groups are repeated in one table. When data from questionnaires and similar devices are analysed the student should be careful than he does not set up the tables incorrectly. For example, it may be reasonable to set up a table as follows:

| | Agree | Neutral | Disagree | Total |
|---|---|---|---|---|
| Item X | 140 | 190 | 170 | 500 |
| Item Y | 180 | 150 | 170 | 500 |

However, $X^2$ contingency test should not be performed on the basis of this table since it is not really a contingency table. Note that the same include only 500 students. Yet this table shows a table of 1000. This is so because each student is classified twice in the table.

## Model Questions

1. Describe the procedure of testing a hypothesis.
2. Write a note on sampling the attributes.
3. Describe the tests for small samples.

# LESSON - 12
# STATISTICAL TREATMENT OF DATA - I

## I. WAGES

In the case of wage a large variety of data are compiled. The following is a brief account of the figures which are collected by the Labour Bureau, Government of India.

i) **Statistics of average wage:** The tables which are compiled show variously the average daily employment, the total wage bill and average annual money earnings of factory workers drawing less than Rs.200 per month. Figures are given state-wise.

In another table, besides the above information a classification of wage earners is further made into those earning less than Rs.200 those earning between Rs.200 and Rs.400 and those earning more than Rs.400 per month. Figures are given state-wise.

A separate table shows the index number of money earnings of factory workers. Figures are given state-wise.

In other tables, the average annual money earnings of factory workers cover those workers who are drawing less than Rs.200 per month and those drawing less than Rs.400 per month.

The indices of factory worker's earnings are also given along with regional index numbers. Index numbers are given industry wise and in another table region wise, classified into six regions as follows:

| | | | |
|---|---|---|---|
| I | North India | IV | West India |
| II | East India | V | Central India |
| III | South India | VI | North-West India |

Figures are compiled under the Payment of Wages Act 1936 and cover only those factories which come under the purview of the Factories Act. The incidence of non-response is not much, because reminders are given and all efforts are made to collect the data where data is not available even after the persistant efforts, threats of cancellation of licence etc. yield the desired results. Figures are separately collected for total wage bill before deduction and the corresponding average daily number of employes and by dividing the former by this latter, average earning per capita are obtained. Earnings include basic wages, allowances, annual bonus and cash value of concessions, we have seen earlier that the geographical coverage of labour statistics has changed from time to time. In the statistics discussed earlier, railways, workshops and seasonal industries (food, beverages, tobacco and gins) are excluded.

ii) **Statistics of average wages in mining industries:** These statistics give average weekly earnings in all local mines, average daily earnings in Bihar mica mines and average daily earnings in all manganese mines. In another table, indices of money earnings of mining workers in certain mining areas are given.

In a separate table wage statistics for coal mines and cotton mills are given for Jharia and Ranigunj (in case of coal mines) and for Ahmedabad, Bangalore, Bombay, Coimbatore and Madras, Delhi, Indore, Kanpur and West Bengal (in case of cotton mills).

The source of the earning and indices statistics is from the Chief Inspector of Mines. The total wages are arrives at after aggregating basic wages, dearness allowances and other cash payments.

**iii) Statistics of average wages of plantation workers:** Data relating to earnings of plantation workers in Assam Valley and Cachar areas are compiled by the Controller of Emigrant Labour and published in the Annual Reports on the working of the Tea Districts Emigrants Labour Act 1932. Figures show the average monthly cash earnings of settled labour per worker. Earnings are given separately for men, women and children for two areas. Figures upto 1955-56 are shown in rupees, annas and pies and thereafter in rupees and paise.

The scope of statistics extends to those tea estates which are submitting returns. Their number varies from year to year. Figures prior to 1953-54 are based an average of March and September while those relating to the years afterwards are based on January to December averages.

**iv) Statistics of Average Wages of Casual Agricultural Labour:** Figures of average daily wages of casual agricultural labour are available for Zonal councils and for the states. Classification of labour is made sex-wise and figures are given as daily averages in paise for workers engaged in various types of agricultural operations.

**v) Statistics of Minimum Wages:** The Minimum Wages Act of 1948 requires the fixation of wages for the lowest paid unskilled workers by the Central and State Governments. Consequently these governments fix the rates of minimum wages of various employments. So far only 33 categories of employment have been taken within the purview of the Act and in which wages have been laid down either by the Central or by the State Governments, for employees working in these employments. The Governments concerned are authorised to extend the application of the Act to other employments also. An employee has been defined as a person 'employed for hire as reward to do any work, skilled or unskilled manual or clerical". Minimum wages, however, are not required to be fixed for employments in which the number of persons working is less than 1,000. For certain employments wages are fixed per day, while for certain others monthly wages are laid down.

The data are initially compiled by the Chief Labour Commissioner and the State Governments and then they are passed on to the Labour Bureau, Ministry of Labour, Government of India, for publication.

## Deflation Time - Series Data of Wages

Deflating means making allowance for the effect of change in price levels. It is the technique of converting a series of value calculated at current prices into a series at constant prices of a given year. In other words, the process of removing the effects of price changes from the current money values is called deflation. Through this process the real value of the phenomenon is

calculated which is not influenced by price changes. Often in calculation of real national income and other economic variables the technique of deflating is used.

The relevant price index number is the deflator, whether it is to be the wholesale price index or consumers prices index or one calculated for specific purpose, would depend on the circumstances of each case. Normally, separate price deflators are found out for deflating the national income data from different sectors of the economy considering the changes in prices in those sectors. For example the general price rise in agricultural products may be more than in the industrial products. The method of deflating is quite simple. Symbolically.

$$\text{Deflated value} = \frac{\text{Current Value}}{\text{Deflator}} \times 100$$

(price index of the year is the deflator)

Note: The idea of "The Purchasing Power of Money" as a measure of real wages or incomes is obtained on deflating the wage series by dividing each item by an appropriate price index eg. the cost of living index. More precisely.

$$\text{Real wage} = \frac{\text{Money Wage}}{\text{Price Index}} \times 100$$

The real wages so obtained may be converted into index numbers as follows:

Real Wages or Income Index Number

$$= \frac{\text{Real Wages of current year}}{\text{Real wage of base year}} \times 100$$

$$= \frac{\text{Index of money wages}}{\text{Price Index Numbers}} \times 100$$

## Example: I

Given the index of money earnings of factory workers in India and all India consumer price index numbers, calculate the index of real earnings and comment on your findings.

| Year | Index of Money earnings | All India Consumer Price index Numbers |
|---|---|---|
| 1986 | 100.0 | 100.0 |
| 1987 | 107.1 | 98.1 |
| 1988 | 107.7 | 96.2 |
| 1989 | 120.8 | 105.7 |
| 1990 | 131.4 | 118.1 |

**Solution**

## CALCULATIONS FOR INDEX OF REAL EARNINGS

| Year | Index of money earnings | All India consumer price index numbers | Index of Real earnings |
|---|---|---|---|
| (1) | (2) | (3) | (4) |
| 1986 | 100.0 | 100.0 | $\frac{100}{100} \times 100 = 100.00$ |
| 1987 | 107.1 | 98.1 | $\frac{107.1}{98.1} \times 100 = 109.17$ |
| 1988 | 107.7 | 96.2 | $\frac{107.7}{96.2} \times 100 = 111.95$ |
| 1989 | 120.8 | 105.7 | $\frac{120.8}{105.7} \times 100 = 114.28$ |
| 1990 | 131.4 | 118.1 | $\frac{131.4}{118.1} \times 100 = 111.26$ |

**Example : 2**

The following table gives the money wages and cost of living index number based on 1980.

| Year | 1980 | 1981 | 1982 | 1983 | 1984 | 1985 | 1986 |
|---|---|---|---|---|---|---|---|
| Wages (Rs) | 65 | 70 | 75 | 80 | 90 | 100 | 120 |
| Cost of Living Index Number | 100 | 110 | 120 | 130 | 150 | 200 | 250 |

Solution

## CALCULATION OF REAL WAGES

| Year | Wages (in Rs) | Cost of Living Index Number 1980-100 | Deflated or Real Wages col 2 / col (3) |
|---|---|---|---|
| (1) | (2) | (3) | (4) |
| 1980 | 65 | 100 | 65 |
| 1981 | 70 | 110 | $\frac{70}{110} \times 100 = 63.64$ |
| 1982 | 75 | 120 | $\frac{75}{120} \times 100 = 62.50$ |
| 1983 | 80 | 130 | $\frac{80}{130} \times 100 = 61.54$ |
| 1984 | 90 | 150 | $\frac{90}{150} \times 100 = 60.00$ |
| 1985 | 100 | 200 | $\frac{100}{200} \times 100 = 50.00$ |
| 1986 | 120 | 250 | $\frac{120}{250} \times 100 = 48.00$ |

**Example : 3**

Following table gives the annual wages of a worker together with the Price Index Numbers. Compute the Index Numbers of Real Income and interpret them.

| Year | Wages | Price index |
|---|---|---|
| 1984 | 200 | 100 |
| 1985 | 240 | 160 |
| 1986 | 350 | 280 |
| 1987 | 360 | 290 |
| 1988 | 360 | 300 |
| 1989 | 370 | 320 |
| 1990 | 375 | 330 |

Solution

### COMPUTATION OF INDEX NUMBERS OF REAL INCOME

| Year | Wages (in Rs) | Price Index No. | Real Wages | Wage Index |
|---|---|---|---|---|
| (1) | (2) | (3) | (4) | (5) |
| 1984 | 200 | 100 | $\frac{200}{100} \times 100 = 200$ | 100 |
| 1985 | 240 | 160 | $\frac{240}{160} \times 100 = 150$ | $\frac{150}{200} \times 100 = 75$ |
| 1986 | 350 | 200 | $\frac{350}{280} \times 100 = 125$ | $\frac{125}{200} \times 100 = 62.5$ |
| 1987 | 360 | 290 | $\frac{360}{290} \times 100 = 124$ | $\frac{124}{200} \times 100 = 62$ |
| 1988 | 360 | 300 | $\frac{360}{300} \times 100 = 120$ | $\frac{120}{200} \times 100 = 60$ |
| 1989 | 370 | 320 | $\frac{370}{320} \times 100 = 116$ | $\frac{116}{200} \times 100 = 58$ |
| 1990 | 375 | 330 | $\frac{375}{330} \times 100 = 114$ | $\frac{114}{200} \times 100 = 57$ |

The index number of real wages has fallen from 100 in 1984, to 57 in 1990. It shows that though the money wage has increased from Rs.200 in 1984 to Rs.375 in 1990, the worker is not better off.

## 2. CONSUMER PRICE INDEX NUMBERS

### Scope and concept of consumer price indices (C.P.I)

The expression "cost of living index" is synonymous with "retail price index", or "consumer price index" and there is little or no difference in their connotation, significance, scope and meaning. These terms are used variously in various countries. It may, however, be noted here that the

purpose of these index numbers is to measure changes in levels of retail prices of goods and services bought and not in levels of prices and standard of living together. In view of this, the Sixth International Conference of Labour Statisticians recommended that "the term cost of living index should be replaced in appropriate circumstances by the term 'price-of-living-index', 'cost of living price index' or 'consumer price index'".

In will be useful if the following scope of a particular cost of living index is specified:

(1) The groups of population covered - e.g. working class wage earners, low-salaried workers, middle class, civil servants etc.

(2) The geographical areas covered - eg. urban areas, areas of a given size - group, a particular city, cities etc.

It is necessary to define the scope of the indices in the manner indicated above because two indices representing changes in the cost of living of two population groups or two geographical areas cannot be used to compare actual living costs of these population groups and geographical areas. The only purpose which those indices serve is to find out changes in the level costs of living as compared to a base period.

The Labour Bureau in their publication, A Guide to Consumer Price Index Numbers have explained the meaning of Consumer Price Index Numbers in the following words: "The consumer price index number is intended to show over time the average change in prices paid by the consumers belonging to the population group proposed to be covered by the index for a fixed list of goods and services consumed by them."

## Method of construction

The CPI is based on a statistical formula which puts together all the items commonly bought by a class of workers and their families as determined from a survey of family and individual spendings.

## Selections of items

The selection of items to form the basis of the construction of this index is not an easy matter. It depends upon the purpose of the index number. Family budget enquiries have to be conducted to find out the pattern of family consumption in various population groups which form the scope of the index. These items of expenditure are of a large variety and vary from place to place and population group to population group. For the sake of convenience they are classified into groups. The Seventh International Conference of Labour Statisticians while considering the analysis of the results of the family living studies, recommended that "for purposes of international comparison the classification of consumption expenditure should be made in such a way as to make possible a grouping or regrouping of items in the following groups and sub-groups:

(a) a group of items, including as separate items, food consumed away from home and alcoholic beverages;

(b) a group of housing, including as separate items, rent, fuel and light and household furnishings and appliances;

(c) a group of clothing;

(d) a group of miscellaneous, including as separate items, the following ten sub-groups: medical care, personal care insurance and other contribution, education and reading, telephone and postage, recreation, transport, tobacco, gifts and charities and servant."

The conference also pointed out that the items of non-consumption eg. income and other personal taxes, interest on personal debts etc, should not be included under consumption expenditure because it is not customary to include non-consumption items in cost of living indices. Similarly, savings in the form of bonds, annuities, bank-savings, insurance premia etc, are also excluded from cost of living indices as it cannot be ascertained when and for what purpose they will utilised afterwards.

While the Seventh International Conference has given, for the sake of uniformity, a classification of items which should be taken into account in the e cost of living indices, in actual practice, these items differ from country to country. In the very nature of things these items cannot be uniform. Even in the case of the same country the details of these items must differ from centre to centre.

## System of weighting

The system of weighting adopted, or the "weighting diagram" as it is called, is an important consideration in the construction of index numbers of cost of living. Index weights are also derived from family budget enquiries. The system is to assign "weights" in proportion to their relative importance as indicated by figures of either expenditure or quantity of consumption in the average budget. The consumption pattern which enables the weights to be assigned has to be examined with reference to the population group to which, the index relates. In specific cases, top suit certain special purposes, assumed weights have to be assigned on theoretical hypotheses. These hypotheses may, for example, be estimated consumption pattern based on secondary data relating to production, export, import, retail trade etc.

## Obtaining values of expenditure

It is generally accepted that the enquiries should cover a period of one year to allow for coverage of seasonal expenses. For non-recurring items such as clothing etc. information for whole year should be obtained. For others four-quarterly data is enough, and families need not be harassed by asking for information at more frequent intervals.

The values of expenditure items, or retail price quotations obviously relate to the period of the family budget enquiries. It is therefore, in the fitness of things that quotations of prices must also be obtained simultaneously with the family budget enquiries. But, as we have seen above, the selection of items can be finished only after the family budget enquiries are completed and analysed. A tentative list of items is prepared and price quotations obtained. The following points

are borne in mind while collecting price quotations:

1. The retail prices should relate to a fixed list of items and for each item, the quality as fixed by means of suitable specification.

2. Retail prices should be those charged to consumers inclusive of taxes for actual cash sales on fixed days of the weeks. Prices charged for soiled, stale, damaged or otherwise imperfect goods should be ignored.

3. "Sale" 'cut' or reduced prices should be mentioned where they are charged.

4. Discount, if automatically allowed, should be taken into account.

5. In a period of price control, where illegal or blackmarket prices are being openly charged, they should be mentioned along with controlled prices. In the case of dwellings, however, the practice is to take an adequate and changing sample of dwellings at each rent collection so that new houses as well as old and houses at controlled and uncontrolled rents are represented in the index.

## Sample family budget enquiries

It is not possible to extend the scope of the family enquiries to a census type of study. The United Nations Statistical Commission also subscribe to the same view. Therefore, it is of vital importance to adopt sound sampling methods in the field of cost of living indices. In this connection it is important that sample should be chosen in such a manner that might enable the calculation or probable error and introduce only the smallest possible amount of bias. The necessity of carrying out these enquiries under the guidance of competent mathematical statisticians cannot be over emphasised. The procedure of selecting the sampled family, enumeration, and tabulation of data must be scientifically done so that evaluation of sampling error is possible.

## Agency of collection

The primary agency through which price quotations are collected is the most important consideration. Price quotations are obtained usually by personal appointed particularly for the purpose; but in certain special cases quotations are also obtained by means of mailed questionnaires and through published price lists.

It will be observed here that the former method. viz.. collection of prices through agents appointed for the special task of obtaining quotations, is the most satisfactory as these agents can be relied upon to visit retail shops and carry out special instructions regarding quality, price variations etc. If however, informants do not require to be explained special points and they can be expected to supply data accurately. Mailing questionnaires are quite suitable. With regard to the agency of collection, the Sixth International Conference of Labour Statisticians made the following recommendations:

"(a) The effective administration of the price collecting programme required the careful selection

and training of price collection personnel.

"(b) Such use of appropriate methods of price variation, such as check pricing, in which price quotations are verified by means of duplicate prices obtained by various agents "or purchase checking" in which actual purchases of the goods prices are made, is recommended".

## System of averaging

After the price quotations have been obtained from retail sources, an average price for each of the items included in the index has to be worked out. Averages are first calculated for the base period of index and later for the month, if monthly indices are published. It is emphasised that the method of averaging should be such as to yield unbiased estimates of average of prices being actually paid by the population as a whole. The problem of averaging is complicated by the introduction of black market or illegal price quotations.

## Construction of Index

After the price quotations have been obtained, the actual construction of the index is a simple matter. An accepted formula, the Laspeyres formula, is most commonly used in which prices are weighted by the quantities of the base year. The formula is as follows:

$$I_n = \frac{\Sigma P_n q_o}{\Sigma P_o q_o} \times 100 \qquad \ldots\ldots\ldots\ldots \quad (1)$$

Where $I_n$ = The index;

$P_n$ = Current prices;

$P_o$ = base prices; and

$q_o$ = base quantities.

The above formula can also be expressed as follows to yield the weighted average of price relatives an fixed base:

$$I_n = \frac{\Sigma P_o q_o \frac{P_n}{P_o}}{\Sigma P_o q_o} = \times 100 \qquad \ldots\ldots\ldots\ldots \quad (2)$$

If, however a weighted average of price relatives on chain base is required, the above formula is modified as follows:

$$I_n = \frac{P_{n-1} q_o \frac{P_n}{P_{n-1}}}{\Sigma P_{n-1} q_o} = \qquad \ldots\ldots\ldots\ldots \quad (3)$$

Any one of the three formulae described is used in actual practice. Sometimes a combination of these is also used in construction of index numbers.

Cost of living index numbers are customarily constructed in stages:

(i) Construction of group indices by any of the three formulae for different groups of items.

(ii) Group indices obtained by (i) above are usually combined by formula no.2 described above, to yield the general cost of living index number.

The construction of cost of living index numbers is not an easy matter and several difficulties have to be faced. The greatest difficulty in the construction of index numbers showing a change in the cost of living of people over a period of time is the human element that is involved. Index measures the changes in the cost of goods and services of constant quantities and qualities between two points of time. But the market conditions and preferences of the consumers do not remain the same with the result that a comparison on the basis of standard or constant quality and quantity is not really possible.

1. As the prices change, the quantity demanded also undergoes a change, however insignificant it may be, change in the quantity demanded must effect the weighting diagram but it will involve fresh family budget enquiries which is not possible at such frequent intervals as the change in prices and the consequent changes in demand. In this connection, the observations of the Sixth International Conference of labour Statisticians are worthy of mention. They recommended that weighting diagram must be adjusted. if necessary, but not at intervals of less than ten years. They also made a very valuable suggestion that small sample studies of consumers purchases must be made in between more comprehensive surveys, to study sample studies of consumers purchases must be made in between more comprehensive surveys, to study change in the preference of buyers and their effect on the weighting diagram. In the case of changes forced upon the consumers by rationing etc the conference recommended a revision of the weighting diagram.

2. The other problem which has to be faced while constructing index numbers of cost of living is the frequent changes in the quality of goods. If this is to reflect perfectly the changes in the cost of living, it must take into account the changes of quality. The expedients followed are: (1) if prices over a period of time available both for new as well as for old qualities, the prices of quality are linked on the basis of the trends of the prices of the new quality. This only means that the basis of the price level of the old quality at the point of change as compared to its prices at the base period. This is the method which is followed in substituting items, as the index level is not affected because of the mere fact of substitution. (ii) The price of the new quality is found out as it was on the base period and the new quality is priced against as base price (iii) Quality difference is evaluated in terms of prices in consultation with retailers, manufacturers, specialists etc and only that part of the difference between the quotations for old and new qualities which represents real price movements taken into account.

It will, however, be seen that in the course of the last few years, technological advances have taken place in such a manner that it is no longer possible to take adequate account of quality changes. Hence no allowance is often made for quality changes.

3. Inclusion of seasonal items is another problem in constructing cost of living index numbers. The Sixth International Conference of Labour Statisticians made the following recommendation in this regard: "(a) Every effort should be made to include a suitable representation of fresh, canned, dried and frozen fruits and vegetables in the list of goods priced. (b) If items cannot be priced every month of the year or if price differences between different seasons are substantial, it is advisable to use incorrected price figures and / or constant weights all through the year (c) Whenever necessary and feasible, account should, therefore, be taken in the methods adopted to the seasonal factors in consumption and prices (d) Provisions should also be made for taking account of seasonal factors in purchase of clothing, fuel and other items showing seasonal variations".

In this connection it is interesting to know that the conference of Commonwealth Statisticians recognised that circumstances might prevent the inclusion of seasonal items, for example, where short term fluctuations in supplies and prices are very erratic. They also recommended a system of adjustment for considerable variations from month-to-month consumption of these items. The Conference emphasised the need for adhering to the concept of an index, ie. a study of change in the prices of a fixed expenditure pattern in the base period. The technique suggested was that vegetable and fresh fruits group must have a fixed group weight within which necessary adjustments may be made in such a way that the group weight is not thereby affected.

## Errors in the Consumer Price Index Numbers

There are chances of grave errors in the construction of the cost of living indices. The very nature of these indices is such that makes them amenable to errors from a large variety of sources. Errors emerge from the following:

### (i) Use of sampling method

The family budget enquiries have to be carried out on a system of sample households. These should be representative of the tastes habits and customs of the population group as a whole. Then the price quotations have to be obtained from sample localities or markets and these should, again be representative of the prices rating in the country.

### (ii) Indifference of the people

Family budget enquiries are the basis of the selection of items and weights. Mostly people are not alive to the uses of such indices and enquiries and they treat the enumerators with contempt.

### (iii) Human element involved

As we have seen earlier, human element is responsible for grave confusion in the construction of these indices.

### (iv) Defects of primary reporting agency

The agency for collection is inefficient and also shrinks when responsibility is entrusted to it. The family budget enquiries and the collection or prices call for considerable patience on the part of the enumerators.

### (v) Credit Purchase

Labourers make most of their purchases in credit, and it is very difficult to find the difference between the real prices prevailing in the market and the credit prices.

## Uses of Consumer Price Index Numbers

Cost of living index numbers are an index of changes in the tastes, habits and fashions of the people of that particular class to which they relate. They provide a reliable study of changes in the consumption pattern of population and are, thus, useful to trade, commerce and industry. They are also useful to industry in another important respect in as much as they provide a satisfactory and reliable basis of wage studies and remunerations. It is on the basis of information available from these indices that policies relating to general economy, wages, prices, taxation, rent control etc, are laid down and wage disputes settled. They also provide an index of the purchasing power of currency.

The Labour Bureau, while commenting upon the use of consumer price index numbers points out that their main use lies in the regulation of dearness allowance of employees to compensate for the rise in prices. If the index with which dearness allowance is linked goes up the employees are given a higher dearness allowance according to an agreed formula. If the index goes down, the allowance is correspondingly reduced.

## Available Consumer Price Index Numbers

The CPI which are available in India relate to working classes only. Efforts are being made to compile index numbers for other classes also. In some cases family budget enquiries have been made, but index numbers are compiled by the Labour Bureau, Ministry of Labour. Government of India and also by some state Governments. These are published under several schemes.

## Labour Bureau indices

The Labour Bureau, Simla brings out two Consumer Price Index Numbers. Their brief details are:

### a) Consumer Price Index Numbers for Industrial Workers

For the propose of this index, items are divided into three major groups i) Food, ii) Clothing and iii) General. The clothing group includes clothing, bedding and footwear. The index is monthly and annual. The annual Index is being taken as the average of monthly Index numbers in its publications. Some confusion is created because the Monthly Abstract of Statistics brought out by the CSO lists 23 centres state-wise, while the RBI lists 21 centres in alphabetical order. The CSO list contains 11 centres which are not to be found in the RBI list while the RBI list contains 9

centres which are not to be found in CSO list. This discrepancy is intriguing particularly because the source (Labour Bureau) is the same for both the lists. Besides these indices, All India Index is also worked out.

Of the various centres included in the CSO list, Doom Dooma, Jalpaiguri and Mundakayam are plantation centres. Jharia and Barbil are mining centres and the others are factory centres.

The weighting diagram is also quite confusing. Different weights have been assigned to groups at different centres. These weights do not total upto 100. The present base year for consumer price for industrial workers is 1982 = 100.

## b) Consumer Price Index Numbers for Agricultural Labourers

For the purpose of this index, items are classified into two major groups (i) Food and (ii) General. The base year is the agricultural year. July to June. The index is monthly and annual, the annual index being taken as the average of monthly indices from July to June.

The weighting diagram is confusing. Different weights have been assigned to different groups in different states. Figures are given statewise. Figures for Haryana, Delhi and Himachal Pradesh are included in Punjab and for Manipur, Meghalaya and Tripura in Assam.

iii) Consumer price index number for working class for certain centres and all-India index base shifted to 1951-100 except where otherwise indicated the index is published jointly giving the a) Labour Bureau series and b) State series. The present base year is 1986 - 87 = 100.

## Construction of Labour Bureau Indices selection of items

Items for construction of the index numbers discussed above have been classified into three broad categories. These are: 1) food 2) clothing and 3) miscellaneous. At some centres the groups are slightly changed to suit local conditions. The items falling under each group are different from centre to centre depending upon the popularity of items at different centres. This is only to be expected in such cases where the selection of items depends upon the taste, fashions and customs of the people. At several centres the group clothing, bedding and footwear is described as "clothing" only.

Items have been selected after careful family budget enquiries. The period of these enquiries is not uniform from centre to centre. Similarly, the scope of these enquiries (eg. factory workers, mining workers or plantation workers etc) is also different at different centres. The sampling frame and the sampling fraction is also different from centre to centre. No comparison between the changes in the consumer prices is therefore possible, under these circumstances.

## Obtaining the price quotations

The price values or quotations are obtained generally with a frequency of one week. The days of the work to which the prices actually relate are different from centre to centre; so also is the agency of collection. In case the index is monthly, a data is fixed for obtaining the price quotations. A uniform practice is not followed with regard to frequency for each group at one centre. For

example, at some centres the quotations are obtained weekly for some groups, and for some other groups the quotations are obtained monthly. The treatment of house rent is also not uniform. At some places the old quotation of house rent is taken while at others it has been revised while still at some others the practice is to revise it at regular intervals, eg. every half year.

The agency of collection differs from centre to centre. The defect of the primary reporting agency which was in vogue sometimes back is only too glaring. Most of the quotations are now obtained through the Economic Intelligence Inspectors or the Marketing Inspectors. But collection of quotation by the way side Railway Station Masters or the clerks of some government offices, or the Factory Inspectors etc., is not rare. In these cases the reliability of the quotations is doubtful.

The quotations are the actual retail prices ruling in the market on the day fixed for collection. Only such markets are selected which are patronised at the centre by the groups of workers to which the index relates.

## System of Weighting and average used

Weighted arithmetic average is used in finding the average at the time of final construction of the index. Weights are assigned to individual items and to groups with the help of family budget enquiries conducted in some cases during the war years and in some cases during the post-war years. Index numbers are usually compiled in two stages; first, group indices are compiled and then they are combined into the general index.

## Base Year

The base year differs form centre to centre, but in most cases now the indices have been brought to a common base so that comparison may be facilitated. In certain cases indices have been worked back to a changed base, but in most cases, there is no continuity between the old and the new base.

The noteworthy thing in connection with the weights, however, is that at most centres the base year for the weights is different than the base year prices. From a technical point of expediency it can be accepted. The point is that frequent family budget enquiries are not possible while, if after a lapse of time some year is considered more suitable to be the base year, it should certainly be adopted in order to make the index more representative of the changes.

## Characteristics of consumer price index numbers

The Labour Bureau's consumer price index number is a statistical measure. It, therefore, suffers from all the limitations and has all those characteristics which are possessed by statistics in general. These indices are compiled for certain specific purposes which are predetermined. They are true on an average under certain given set of conditions. They are capable of misinterpretation if the method and technique of their compilation and the type of data that enter into their construction are not carefully assessed. The following characteristic may be noted in order to understand their true significance:

1) The consumer price indices compiled by the Labour Bureau relates to a particular population group and geographical area. They are not applicable to the population as a whole but to that group of population for whose specific purpose they are constructed.

2) While these indices are average themselves, they are true on an average only because within a population group also there are wide variations with regard to tastes, customs and fashions and the consumption habits differ from family to family in each group. Only the average experience of individual families is reflected in these indices.

3) Considerable amount of approximation and estimation enters into the construction of these indices. This is because the budget of a representative family is derived from the study of a sample of budgets and a selected list of items is taken out of these sample family budget enquiries. Weights are greatly approximated and price quotations are also estimated, because they are taken only for fixed days of the week, month, year etc. It is a different thing that these errors which tend to creep in are cancelled out and the magnitude of final error is very small.

4) The index is a price index and not a cost of living index. It will be observed that in India formerly these indices were referred to as cost of living indices. But because of the confusion which they were likely to create it was deemed the former nomenclature was likely to create an impression that these indices were designed to measure standard of living, rather than changes in the price levels of commodities which are commonly consumed by people. Cost of living is a phenomenon which is governed by a number of factors such as income tastes quality and quantity of goods and services consumed etc. The consumer price index tells us how much it costs to maintain the same level and type of consumption under changed price levels are compared to some previous data which may have been taken as the base.

5) The consumer price index does not take into account changes in expenditure pattern, tastes, habits etc of the people. It is obvious from the above point that it is not a measure of changes in the tastes, habits and customs of people. By taking one year as the base and by selecting a set of commodities for all times to come in the life time of a particular index, the index precludes all possibility of being regarded as an indicator of changes in the consumption habits of the people.

6) The index measures relative changes in prices and not absolute changes. The index is a relative measure by its very nature. It is not capable of taking into account the changes in absolute terms but only relative changes. This means that proportionate changes in prices of various commodities are reflected in the index.

## Criticism of the consumer price index numbers

The object of compiling the consumer indices is to measure changes in the cost of living of various classes of people. The indices available in India relate mostly to the working classes. In a few cases index numbers for other classes like agricultural Labourers and the employees of the Central and State Governments have also been collected. Before going into details of the short-comings of these index numbers it must be realised that index numbers are studies in percentages

and averages. They can give an idea only of average changes in the cost of living, not of changes in the cost of living of individual families. It must further be realised that the technique of construction of the indices viz. the selection of items, the base year and the weight are decided upon as a result of the family budget enquiries which cannot be conducted as an annual feature. Therefore, the interpretation of these indices is not applicable for a long period of time. The consumer price indices as published in India, however, suffer from certain drawbacks which tend to undermine the utility of these index numbers.

### i) The items

The items on which the indices are based differ from index to index. This is only natural because the indices are intended to be regional in their nature. However, certain items which should be the feature of all indices seem to have been left out of some index numbers. For eg. the Jharia index does not include fuel and lighting. Similarly house rent is not included in Bihar. The miscellaneous group has been dropped in Bihar and Orissa while in Tamilnadu, this group has been made prominent by including nine items, under this group. Some vital items of the daily necessity of workers have been excluded from the list, eg. footwear, underwear, umbrella, blankets etc. But the family budget enquiries are themselves very old and in the meantime the pattern of consumption of the people has undergone a change. The tastes and fashions of the people have a direct bearing upon the index number. The workers now do not consume only one cup of tea per day, or do not smoke only five bidis in a day, or shave only once a month. The frequency is now larger for these items. Similarly, they do not consumer broken or Patni rice now, but going for whole and higher quality rice. Again, certain varieties of cloth which used to be consumed in 1939 when family budget enquiries were held in Bombay and Kanpur are not produced at all. The inclusion of such items tends to distort the index number.

### ii) The base year

The following are accepted as a common base for purposes of comparison in all economic studies

1. Consumer Price (All India) Industrial Workers - 1982 = 100
2. Urban non-manual employees (1984 - 85) = 100
3. Agricultural Labourers - (1986-87) = 100

### iii) The average and the system of weighting

As has been observed earlier, the average used in the construction of these indices is weighted arithmetic average. From a technical point of view, geometric mean is better as it makes the index reversible and gives equal weight to equal ratios of changes. Thus, it allows the shifting of the base and the change in the items without distorting the index. But geometric mean is not easily understood and the difference between an index on geometric mean is very little; hence arithmetic average is considered to be better or balanced. However, with regard to weights an observation may be made here. The wieghts differ from place to place. They were based on family budget enquiries held far back and are quite out of date in the present context.

It is clear that consumer price indices which are being constructed now are out of date and a revision is overdue. The coverage of the index numbers is very limited. At present it extends to

the industrial workers only. In a few cases some other categories of workers have also been included but that has not materially acted the situation. New family budget enquiries are overdue and must be made without further delay. The Government of India have already started working on new schemes and the work is well under way.

## New schemes of Consumer Price Index

The Labour Bureau which brings out the Consumer Price Index every month will now have 1980-81 as the new base. Formerly, the base was 1960. The new index will have hundred points. The consumption pattern has undergone a change during the last twenty years and the base as also the items have been rendered unrealistic. Therefore, acting up on the recommendations of the I.L.O., the changes have been introduced. So far (July 1981) besides fixing the base year as 1980-81 no other details have been made available to the people at large.

With this exercise which is going on at the Labour Bureau level, another allied activity is going on which deserves mention in the context of price statistics. The Bureau is planning a detailed survey with the ultimate object of monitoring the supply of essential goods at fair prices in the country, particular, in the far fund areas. This survey would be in addition to the two surveys already being made for industrial workers and farm labourers.

# 3. EMPLOYMENT

A large variety of data relating to employment are compiled by the Labour Bureau. They relate to those factories which submit returns, while in the case of those not submitting returns estimates have been made on the basis of inspection reports, previous year's data and applications for licences and registration. The scope of these data extend only to such factories which come within the Indian factories Act, viz, those factories which employ 10 or more workers and use power, those which employ 20 or more persons and those which have been specially brought under the Act by an order of the state Governments concerned. The data cover persons who are employed by the factories for any purpose on wages or otherwise and include clerical staff and supervisory personnel also. The integration of states in 1949 and reorganisation in 1956 has changed the coverage of the data rendering them incomparable.

The following figures relating to employment are available.

### i) Average daily employment in factories

The figures are given state-wise as well as industry-wise for states and for the country as a whole and are collected under AŞI.

### ii) Number of working factories and number of factories submitting returns

Here again the figures are given state-wise and all India figures are also given.

### iii) Average daily employment in mines.

The figures relating to mining industry are given separately are classified as follows: (a) coal (b) mica (c) manganese and iron ore, (d) others.

In the case of coal mines employment figures are further classified into three categories viz, below ground, open cast workings and above ground. Mining statisitcs are compiled by the Directorate General of Mines safety, Ministry of Labour. The geographical coverage of the data has not been uniform. Upto 1950 they extended to part A states and Ajmer and Delhi and thereafter they relate to the country as a whole. Mines and quarries have also not been covered uniformly by these statistics because in 1955 a number of them were brought within their purview by a notification of the Government of India.

## iv) Average Daily Employment in Plantations

Labour engaged in (a) tea (b) coffee and (c) rubber plantations is covered by these statistics. Figures of daily employment as well as indices with base 1951 are given in the same table. Data are compiled by the Directorate of Economics and Statistics, Ministry of Food and Agriculture. Base years are subject to change.

## v) Employment in Railways and in Posts and Telegraph

Figures for labour including gazetted officers and sub-ordinate employees in offices, open lines and construction, engaged in railways are given. For posts and telegraph, figures are given separately from railways and include non-gazetted staff only. Side by side, index numbers of employment are also given with 1951 as the base Data are supplied by the Railway Board (Annual Report) and by the Director General of Posts and Telegraph. Base years are changed to suit the situation.

## vii) Employment in Shops and Commercial Establishments

The shops, commercial establishments and restaurants, theatres etc are shown side by side with the number of persons working in these three categories of establishment along with their total. Figure relate to those areas only where the shops and Commercial Establishment Act applies. Arrangements for the collection of data is very unsatisfactory and returns are generally not supplied.

## viii) Employment in Central Government Establishments

The figures do not include such central government establishments for which separate tables are given viz, the Railways, Armed Forces, Embassies and the Missions abroad. The classification of this data as made as follows.

a) Administrative and executive

b) Clerical

c) Skilled and semi-skilled and

d) Unskilled

## ix) Estimated Employment in Public and Private Sector

Number of workers in the public and private sector, further classified in the case of the public, sector into: a) Central Governments b) State Government c) Quasi-Government and d) local bodies and in the case of the private sector into: a) large and b) small establishments.

Establishments employing 25 or more workers are referred to as 'large' and those employing 10-24 workers are referred to as 'small'. Figures are given year-wise and for each year quarterly figures are also available.

### x) Employment in Major Industries and Services in Public Sector and Private Sector

Figures are given in two separate tables for Public and Private Sectors. Industries are classified into nine classes: a) Agriculture, hunting etc. b) Mining and Quarrying c) Manufacturing d) Electricity, Gas and Water e) Construction f) Wholesale and Retail trade g) Transport, Storage and Communications h) Financing Insurance and Real estates and i) Community, Social and Personnel services. This classification is as per NIC. Till 1974 the data was on SIC.

Data for (ix) and (x) are obtained from Directorate General of Employment and Training, Ministry of Labour.

Detailed statistics concerning the labour force were collected in India at the time of the various censuses. The population was divided into workers and non-workers, further sub-divided into nine occupational categories and eight categories respectively.

### Data Relating to Employment of Agricultural Labour

Statistics pertaining to agricultural labour in this country related to employment and earnings of labourers in the agricultural sector and also their mobility. Statistics of mobility of agricultural labourers are conspicuous by their absence and any effort to make a broad guess in this respect might lead to dangerous conclusions. The peculiar characteristics of agricultural labourers is that he is a multi-profession man. He possesses no land on his own, yet he is given some land by his master or persons for whose land he works. Thus he is a cultivator also. Then he carries on his caste profession also i.e. serves as a barber, blacksmith etc. Hence, he is available for agricultural or industrial work only when he is free from his own personal work. Again agriculture is a profession in India in which there is much disguised unemployment. The cultivator is not busy all the year round. Agriculture depends upon monsoon and when it rains the cultivator is free from the botheration of irrigating the land. In off period such cultirvators move to the nearby cities and seek employment either as domestic servants or daily wage earners in building industry. They also work as rickshaw pullers. Hence it is quite clear that our Government has not yet developed any such statistical machinery which might keep correct and up-to-date records of landless agricultural labourers in rural areas.

## 4. PRODUCTION

### Indices of Agricultural Production

Index number of agricultural production are very useful for studying the trend of food supplies in the country, especially at the present time when India is making every possible effort to increase agricultural production and trying to keep pace with increasing population. Although construction of index numbers started in India as early as the last quarter of nineteenth century, agricultural production did not find favour with those who constructed it. Most of the early index numbers were

confined to price of export and import articles and food grains. Such index numbers were compiled with the base year 1873 and published in a quinquencial journal, The index number of Indian Price, which was annual supplements. It was during and after the Second World War that index numbers covered a wide variety of articles and trade in the country and now they are available in respect of a large number of production, consumption and trade items.

At present there are three index numbers of agricultural production in India. They are as follows: (1) Reserve Bank of India (2) FAO Index and (3) Ministry of Food and Agriculture series.

All the above index numbers relate to gross production of crops as published in the estimates of area and production of principal corps in India issued by the DES. These index numbers are weighted index numbers and are based on final estimates of production subject to revision the next year when revised estimates become available. Laspeyer's formula as given below is used for the construction of the above index numbers:

$$P_{01} = \frac{P_1 q_1}{P_0 q_0}$$

where:

$P_{01}$ = Price Index number of current year, without the factor 100, which, if multiplied by 100, will yield the index.

$P_1$ = Current year price

$q_0$ = Base year quantity

$P_0$ = Base year price

Separate index numbers are available for individual commodities and an index is available for all commodities which si known as the General Index. Such index numbers are avaialble annually. Separate figures are also avaialble for foodgrains and non foodgrains.

## 1) Reserve bank of India Index Number of Agricultural Production

This series is compiled by the Reserve Bank of India annually and is available in its monthly bulletin. "Reserve bank of India Bulletin" for the month of December each year. Index numbers are available with the average of 1936-37, 1937-38 and 1938-39 as the base for undivided India from 1939-40 to 1948-49. The base year of the index number has been changed to the year 1948-49. At present RBI series are available with 1981-82 as base year.

## 2) FAO Index Number of Agricultural Production

The Food and Agricultural Organisation of the UNO published a series of index numbers of agricultural production relating to different countries of the world including India. The purpose of FAO index is to make inter country comparisons of agricultural production and with that end in view, it takes into account world price relationship of agricultural commodities in tis weighting system.

The index number is weighted although the system of assigning weights is highly complicated. weights are based on wheat relative prices only because wheat is a world commodity. Commodities included in FAO series of index numbers are available. In addition to the common agricultural commodities, certain others are also included in the list such as wine, fruits, vegetable products and livestock products etc. In the construction of index numbers relating to India such commodities are omitted from the list which are not grown in India.

## 3) Ministry of Agriculture - Index Number of Agricultural Products

The DES used to bring out a series of indices of agricultural production covering 19 commodities. The base year was the quinquenniam ending 1938-39. These index numbers were considered as out of date in view of the progress made in the compilation, coverage and technique of crop estimates. The Directorate is now bringing out a revised series of indices of agricultural production.

### Selection of items

The revised series comprises 28 items for which the estimates of production are readily available in the form of crop estimates. The crops are divided into two main groups and six sub groups. The details are given in the weighting diagram.

The Government has decided that more commodities should be included in the index but since reliable annual estimates of production are not available in respect of the other crops. For the time being is has been decided to confine the index to these 28 items only. As soon as facilities are available more cops will be brought within the coverage of the index. The question of enlarging the coverage is to be reviewed by the Government periodically.

### The base

The original index had the five year period, 1934-35 to 1938-39. In 1945-46 the Government decided to suspend the publication of these index numbers in view of the desirability of shifting the base and extending the coverage of items. Since 1950-51 was chosen as the base several considerations have weighted with the Government in choosing the base year. The base year for the index No. of wholesale prices has been now fixed at 1993-94. The base year for index number of industrial production is also 1993-94.

### System of Weighting

Different crops have been assigned weights in proportion to the total value of production of each crop during the base period. The evaluation of production has been done at the annual harvest prices current during the year or wholesale prices, whichever available.

It has been assumed that the variation in production in the non reporting areas is the same as in the reporting areas in the aggregate. The second assumption is that the variation in the figures of production based on the new method of estimation is same as in the figures based on the old method of estimation. This means that for the transitional year when the method of estimation

undergoes a change, production figures should be made avaialble on the basis of the new as well as the old method of estimation. Allowance have been made for changes in the coverage, such cases being rare. The following table gives the commodities and the weights assigned.

| Commodity Group | Weights | Commodity group | Weights |
|---|---|---|---|
| I. Food grains | | II. Non food grains | |
| (a) Cereals | | (a) Oil seeds: | |
| (i) Kharif: | | Groundnut | 5.7 |
| Rice | 35.3 | Sesamum | 1.2 |
| Jowar | 5.0 | Rapeseed and Mustard | 2.0 |
| Bajra | 2.7 | Linseeds | 0.8 |
| Muize | 2.1 | Castor seed | 0.2 |
| Ragi | 1.2 | Oil seeds | 9.9 |
| Small millets | 1.5 | (b) Fibres | |
| Kharif cereals | 47.8 | Cotton | 2.8 |
| (ii) Rabi | | Jute | 1.4 |
| Wheat | 8.5 | Mesta | 0.3 |
| Barley2.0 | | Fibres | 4.5 |
| Rabi cereals | 10.5 | (c) Plantations Crops | |
| Cereals | 58.3 | Tea | 3.3 |
| (b) Pulses | | Coffee | 0.2 |
| Gram | 3.7 | Rubber | 0.1 |
| Tur | 11.1 | Plantations crops | 3.6 |
| Other pulses | 3.8 | (d) Miscellaneous crops | |
| Pulses | 8.6 | Sugarcane | 8.7 |
| Foodgrains | 66.9 | Tobacco | 1.9 |
| | | Potato | 1.0 |
| | | Pepper (black) | 1.2 |
| | | Chillies (dry) | 2.0 |
| | | Ginger (dry) | 0.3 |
| | | Miscellaneous | 5.1 |
| | | Non Foodgrains | 33.1 |
| | | All commodities | 100.0 |

## Technique of construction

A slightly complex system of construction of the index has been adopted. The index numbers, to begin with, are constructed on a chain base. This is done to provide for the change in the geographical coverage of the estimates and the methods of estimation. Thus, for each crop, the all India production for a year has been expressed as a production relative with the production of the previous year as the base. Subsequently these production related items have been linked to the base period production to yield the production index for that crop. The sub-group, the group and the all commodity index numbers are arrived at by finding the weighting arithmetic average of these crop production indices.

The index numbers of the current year are provisionally based on revised forecast figures of the previous year. Suitable adjustments are made for changes in the geographical coverage. This becomes necessary in view of the fact that the revised estimates of the current year are avaialble only the next year.

Figures of production are those of gross production only. No allowance is made for seed or wastage. It will be noted that the FAO series of Index Numbers makes allowances for crops, milk and skimmed milk etc, utilised as animal feed whether they came from domestic or foreign sources. This is desirable to avoid double counting, if the crop is counted as produce and later the cattle produce is also taken into account. This problem,. however is not relevant in the Indian context as livestock produce is excluded from the scope of this index.

## Indices of Industrial production

Index numbers of industrial production are the best measure of economic progress of any country. The industrial progress of India has been steady for the last one decade as a result the successful execution of the five year plans. A number of indices of industrial production are compiled in India by official and non-official agencies. The general index of industrial production is the most popular of these. The details of the indices of industrial production is the most popular of these. The details of the indices of industrial production are given below:

## 1. Index Numbers of Industrial Production

A number of indices of industrial production have been prepared from time to time. Their coverage, weighting diagram and technique of construction have been changing from time to time. The Index is compiled by the CSO. Presently the base year has been put at 1993-94.

## III. Index Numbers of Mineral Production (Quantity and Value)

The Indian Bureau of Mines compiles index numbers of (a) quantity and (b) value of mineral production.

The quantity index is compiled with the economic activities in the mining and quarrying sector in view. The index takes within its scope 61 minerals classified into four groups as per the NIC - 1970 classification. (i) coal mining (including lignite); (ii) crude, petroleum and natural gas;

(iii) metallic mining and (v) non-metallic mining. The prescribed substances under Atomic Energy Act, 1962 and minor minerals are not included in the index. The base year for the index as also for the weights are changed according to situation.

The weights are in proportion to the net value added for each items to the total value of the entire mineral sector excluding minor minerals and prescribed substances.

The value Index is compiled on the basis of items and weights as for quantity Index. It shows the trend of gross value of production at pit head prices.

## 5. CONSUMPTION

The main source of data on consumption in India is the comprehensive family budget studies undertaken by the National Sample Survey (NSS). The consumer budget inquiry forms an integral and probably the most important part of the general programme of the NSS to conduct national socio-economic inquiries to provide data needed for planning for national development. Since 1951, the data on household consumer expenditure relating to India as a whole have been collected every year, sometimes twice a year through a series of repetitive surveys called Rounds. These Rounds have been conducted on a probability sampling basis in the form of atleast two independent sub-samples.

The important concepts and definitions used in the NSS Rounds are follows.

### Household

A household comprises a group of persons usually living together and taking principal meals from a common kitchen at least for 16 days during the reference period, ie. 30 days preceeding the date of inquiry.

### Household Consumer Expenditure

Household consumer expenditure comprises all expenditures incurred by a household on domestic account during the reference period. This means that the household consumer expenditure is exclusive of all expenses towards the enterprise activities of the household. A uniform concept of consumption is, however, not followed for all items of consumption. Broadly speaking, for expenditure categories like food, intoxicants and fuel and lighting which are used in a single cycle of the consumption process, consumption means actual amount consumed during the reference period. On the other hand durables and semi-durables which will not be used up in a single cycle of the consumption process, consumption implies the amount acquired during the reference period. Again, different components of consumption are not evaluated at the same price. Consumption out of goods purchased on cash is evaluated at the actual purchase price irrespective of the period of purchase, that out of goods received by way of transfer receipts in kind, like loans, gifts, charities, free collections is evaluated at the average retails price prevailing in the locality of the sample household during the reference period and that out of goods produced by the household at the producer's price.

While consumption out of transfer receipts, as mentioned above, forms part of the consumer expenditure, transfer payments of all kinds (loans, gifts, etc) are excluded. Expenditure on purchase and construction of residential houses is considered as expense on capital account and is, therefore excluded from household consumer expenditure; but expenses towards maintenance and repairs of residential buildings are included. In case of other durable and semi-durable items of consumption not only the expenses for their maintenance and repairs but also for their purchase and construction are included in the household consumer expenditure.

## Data Collected and Published

National Sample Survey (NSS) collects detailed and useful information than it has been able to publish. For each narrowly defined commodity entering into a consumer budget, the NSS collects data on consumption out of purchased articles home grown stocks and transfer receipts. These data are recorded in terms of value as well as quantity wherever possible, eg. age, sex composition of members and their occupation are also collected. As against this, the published reports contain the following details for each of the classes, (i) average total monthly expenditure per capita (ii) the average monthly expenditure per capital under 12 groups of commodities (cereals and cereal substitutes, milk and milk products, edible oil, meat, egg and fish, sugar, salt, other food items, clothing, fuel and lighting, rent, taxes and other non food items) (iii) total number of sample households and (iv) average household size. These cross tabulations are presented for at least two independent sub-samples as well as the combined samples separately for rural and urban sectors. The sectoral information is furnished not only at an all India level but also invariably at the regional level. However, regions refer not population zones comprising geographically continuous states before the 13th Round and to states as such from this Round onwards.

## Household Size and Occupation

Data on household characteristics like age, sex, composition and occupation have been collected in the NSS Round since their very inception, yet it is only in their latest report on household consumer expenditure pertaining to the 19th round that consumption data cross classified by income and household size and by income and occupation respectively, have been presented of course only at the all India level, separately for rural and urban communities.

Statistical studies on consumer behaviour are needed for various purposes. Thus, the elasticities of demand are necessary for choosing the appropriate items for indirect taxation as well as for assessing the tax yields and for projecting the aggregate demand for a commodity for studies on levels of living and for econometric studies on the functioning of the economy.

## Model Questions

1. Explain the statistical data available for wages, employment, production and consumption in India.
2. Write a note on the formation of Cost of Living Index.

## LESSON - 13

# STATISTICAL TREATMENT OF DATA - II
# NATIONAL INCOME AND SOCIAL ACCOUNTING

## Definition

National income refers to the net money value of the total economic goods and services produced by nationals of a country during a given period of time. The following are some definitions which exphasise its various modes of calculation, viz., output, income and expenditure:

"A national income estimate measures the volume of commodities and services turned out during a given period counted without duplication".

— National Income Committee

"The national income is the money measure of the aggregate of goods and services according to inhabitants of a country during a year including the net increments to or excluded the net decrements form their individual or collective wealth".

- Bowley and Robertson.

A comprehensive definition has been given by the United Nations which refers to three main methods of calculation as follows:

"an aggregate of net values added in all branches of economic activity during a specified period together with net income from abroad."

"as the sum of expenditure on final consumption of goods and services plus domestic ad foreign net investment".

## Utility of National Income Estimates

The national income estimates are extremely useful in a number of ways:

1. The detailed estimates of national income throw light on the working of the economy. They show the contribution made by the various branches of industry and trade to the national product. If the contribution of any particular sector is less in comparison to the men employed there in, one can investigate the causes responsible for this so that suitable steps could be taken to increase the share of that sector.

2. National income figures are indispensable in economic planning. In fact, in the absence of detailed estimates of national income most of the economic policies would be a leap in the dark. Also a balanced economic developed will not be possible in the absence of national income statistics.

3. National income estimates throw light on the standard of living of the people. If the national income is increasing, other things being equal, the standard of living of the people are improving. Infact, to get an idea as to how standards are changing, the per capital income is computed.

4. Various Government departments need national income statistics to guide their current operations. The Treasury Department, for instance, uses them to estimate future tax receipts. The departments of labour trade, industry, transport, etc., obtain a great deal of information from national income statistics which help in formulating suitable policies.

5. National income data are employed to evaluate the progress and performance of the plans of economic development. A dependable guide would be to see if the targeted rates of growth of national income, per capita income, per capita income, investment, consumption and output have been realised.

6. A comparison of the rates of growth of national income, per capita income, investment etc., can be made over a period of time and also over space. Intersectoral and inter-regional comparisons are of great use in assessing existing policies and formulating new ones.

7. From an economic point of view the inter-regional or inter-state variations in the distribution of income is very important. For balanced economic growth it is necessary that there should be no disparities in the distribution of incomes as among various regions or states. Regional distribution of income gives an idea of the progress of regions over a period of time.

## Methods of Estimating National Income

There are three popularly known methods of calculating national income, namely:

1. The Income Method.
2. The Output Method.
3. The Consumption - Savings Method.

## 1. The Income Method

This method derives the national income as sum of net income received by individuals and business enterprises. It relies primarily upon the large volume of data gathered by income tax authorities and some times upon the special studies relating to earnings of various occupational groups and to their family budgets. The statistical office of the United States recommended the following summary classification of national income by this method.

a) Wages and Salaries:

   Private

   Government civilian

   Military

b) Other labour income.

c) Income of unicorporated enterprises.

d) Corporate profits before taxes.

e) Profits of public enterprises

f) Interest.

g) Net rents on lands and houses.

h) Net interest and dividends and other incomes from abroad.

The sum of the above categories will give national incomes at factor cost. In a country like India where the proportion of tax-payers is very small this method alone will not be suitable for calculating national income because of the inadequacy of coverage.

## 2. The Output or the Product Method

According to this method national income is calculated adding the net values of goods and services in all branches of economic activity together with the net income from abro¬d. It utilizes the large body of the production and trade statistics avaialble in most of the principal countries of the world. If the valuation of the output of the goods and services is done at market prices, the national income is said to be national income at market prices and if the valuation is done at prices which equals the payments received by the various factors of production only, national income is said to be national income at factor cost. The Statistical Office of the United States recommended the following summary classification of national income by this method.

a) Agriculture, forestry, hunting and fishing.

b) Mining and quarrying, manufacturing industry, electricity, gas, water and sanitary services.

c) Building and construction

d) Transportation and communication.

e) Commerce.

f) Banking, Insurance and real estate.

g) Government services.

h) Miscellaneous services.

i) Net income from abroad.

This methods is more widely used in underdeveloped economies but one must guard against the possibilities of omissions and duplication which are greater in this method.,

## Precautions in the use of this method

Great care must be exercised while adopting the product method otherwise the estimate of national income obtained may be highly misleading. Some points worth considering are:

1. National income should always be expressed with reference to particular years and should be stated as at current price and at constant prices separately to avoid confusing the one for the other.

2. In order to avoid double counting only the final products (excluding raw materials and intermediate products) should be taken into account.

3. The method should be applied only to such sectors which give rise to the production of concrete goods such as mining, manufacturing, agriculture etc. It should not be applied to the service sector of the economy, i.e, transport, commerce and communication.

## 3. The Consumption - Savings Method

National income according to this method equals sum of expenditure on final consumption goods and services plus domestic and foreign net investments, symbolically as, Keynes expressed it.

$$Y = C + S$$

where Y = Total Income

C = consumption, and

S = savings.

This method is used less extensively than the other methods because it calls for data not generally avaialble. No country has as yet complete, continuous and reliable series on the volume of consumer expenditure and savings. At present the consumption - savings method is used as a stop-gap when industrial or income statistics are badly Packing.

Theoretically, all the three methods are related and should give identical results. In practice, all the three methods are combined because data regarding any one are not perfect. No doubt this is logically defective and the result cannot be additive and some error is introduced, but this error is much smaller than if only one method based on secondary data is employed.

## Concepts of National Income

The following integral concepts are used in the national income analysis of a nation.

1. Gross National Product (GNP)
2. Net National Product (NNP)
3. National Income (NI)
4. Personal Income (PI)
5. Disposable Personal Income (DP)
6. Discretionary Income
7. Per Capita Income

1. **Gross National Product (GNP):** GNP is the money value of national production of goods and services for a year. It is calculated by avoiding double counting and excluding intermediate products.

2. **Net National Product (NNP):** NNP is the net production of goods and services in a country during the year. It is GNP minus the value of capital consumed or depreciation during the year. NNP is also called National Income at Market prices.

3. **National Income at Factor Cost:** National income at factor cost is NNP minus indirect taxes plus subsidies. It is total of all income payments received by all factors of production.

4. **Personal Income (PI):** Personal income is the actual income received by all individuals in the country from all sources. It is calculated as follows:

| | | | | |
|---|---|---|---|---|
| Personal Income | = | N.I | - | Corporate income tax |
| | | | - | Undistributed corporate profits |
| | | | - | Social security contributions |
| | | | + | Transfer payments. |

5. **Disposable Personal Income (DPI):** This is simply after tax of personal Income D.P.I = PI - (personal taxes, property taxes and insurance payment)

6. **Discretionary Income:** Discretionary income is the actual income at the discretion of the people to spend. Discretionary income = D.P.I. - Previous hire purchase commitments.

7. **Per capita Income:** This is the average income per head. This is calculated by dividing N.I. by the size of population of the country.

## Estimation of National Income in India

The preparation of a consistent and comparable set of annual estimates of national income on an official basis is a recent phenomenon in India. In 1949, the Government of India appointed the National Income Committee (NIC) under the Chairmanship of Professor P.C. Mahalanobis to make recommendations regarding the estimation of national income, the improvement of the statistical data on which the estimates were to be based, and the promotion of research in the field of nationalincome. It submitted the final reporting 1954 which contained a detailed account of the nature of statistical data avaialble, the limitations of the avaialble data gaps in the existing data base, and an outline of suitable methods which can be used for estimating national income in India. Since then the Centra Statistical Organisation (CSO), Government of India, prepares and published on a regular basis the annual estimates of India's national income.

Subsequent to the publication of the Report of the NIC, there were several improvements in the availability of primary statistical, data and also some new studies on the reliability of the data sources and the suitability of the method - adopted for national income estimation to take note of these changes, the CSO brought out a study in 1961 entitled proposals for a Revised Series of National Income Estimates for 1955-56 to 1959-60. It contained an analysis of the improvements considered necessary and possible and invited critical comments on the recommendations made in the study. The discussion stimulated by CSO's publication resulted in another publication in 1967 entitled Brochure on Revised Series of National product for 1960-61 to 1964-65. This

publication discussed the limitations of the estimates prepared before 1967 now known as "Conventional Series" and outlined the revised methodology to be followed by the CSO to construct what is now known as the "Revised Series" of nationalincome estimates in India. The revised methodology outlined in the brochure was formulated in the light of the recent changes in the availability and quality of various types of statistical data related to the Indian economy.

## METHODOLOGY OF ESTIMATION

We will examine briefly the basic methodology of estimation outlined in the brochure. For the purpose of estimating national income in India the various economic activities are classified into the following broad sectors:

1. Agriculture including Animal Husbandry.
2. Forestry and Logging
3. Fishing.
4. Mining and quarrying
5. Manufacturing:
   a) Registered manufacturing
   b) Unregistered manufacturing
6. Construction
7. Electricity, Gas and Water supply
8. Transport, Storage and Communication
9. Trade, Hotels and Restaurants
10. Banking and Insurance
11. Real estate, Ownership of Dwellings and Business Services
12. Public Administration and Defence.
13. Other services.

Depending upon the nature and extent of data available for each of these sectors, various estimation procedures are adopted for different sectors. On the basis of the nature of basic statistical data and the related source material avaialble for the estimation of national income, we can classify these sectors into three broad categories. The first category will consist of six sectors which include first six items listed earlier except unregistered manufacturing. For each of these sectors the basic data on output, input and prices are avaialble on a more or less regular basis.

The second category consists of railways, electricity generation and transmission, communication, organised banking and insurance, real estate and public administration and defence. Fairly comprehensive data on all types of factor incomes are readily available from the published annual accounts of the undertakings in each of these sectors.

The third category includes the remaining sectors for which no satisfactory data is available either on output or input or on different types of factor incomes. Unregistered manufacturing, trade, hotels and restaurants and other services are the major sectors that fall under this category.

Generally for measuring national income the product approach can be directly followed only in the case of commodity producing sectors while the income approach is the only available alternative that can be followed in the case of various types of services for which output cannot be directly measured. Thus it is obvious that in the case of the Indian economy, the product approach can be followed only for the six sectors falling under the first category mentioned earlier, while the income approach in its direct or indirect form has to be followed for deriving estimates for the sectors falling under the remaining two categories.

The estimates relating to the six sectors in the first category are obtained by adopting the following procedure:

**The first step** consists in deriving the gross value of total output in a particular sector by multiplying the annual output of various products with corresponding average prices duly adjusted for trade margins wherever necessary. It may be noted in this connection that for a number of products in some sectors like agriculture, animal husbandry, forestry and fishing, adequate data on output is not avaialble on an annual basis and, therefore, it has to be derived by following the methods of interpolation or extrapolation as the case may be.

**The next step** consists in estimating the total value of the inputs or the intermediate products used in producing the different products. The data avaialble for this purpose is by and large scattered and scanty especially in the case of agriculture, animal husbandary, forestry and fishing. Moreover, it is available only at some given points of time. The annual estimates of the value of inputs are, therefore, derived by suing input-output proportions calculated basically for a few bench mark years for which some information is avaialble.

**Finally,** the estimates of net value added are obtained by deducting the total value of inputs and the depreciation allowance from the corresponding estimates of the gross value of output. The estimates of the value added at the constant base period (1960-61) prices are derived by applying the prices actually prevailing in 1960-61 to the corresponding figures of output and input for different years.

The estimates relating to the sectors falling under the second category are obtained by following the income approach in its direct form. The year to year data on compensation for employees, interest payments, rent operating surplus and depreciation allowance are readily available from the published annual accounts of each of the undertakings in different sectors. They are, therefore, used to estimate the flow of total factor earnings which taken together, constitute the value added in the respective sectors. Evidently, the estimates relating to these sectors can be regarded as fairly satisfactory and reliable especially in relation to the estimates for other sectors. The corresponding estimates at the constant base period (1960-61) prices for different years are obtained by carrying forwards the detailed base year estimates, using appropriate physical indicators representing changes in the real product of the respective sectors under this category.

The estimates relating to the sectors falling under third category form the weakest link in the national income estimates for the Indian economy. Since no systematic data is available for any of these sectors, the estimates are obtained by following a round about method. The method consists of multiplying the estimated average value of output per worker in the respective sectors. The estimates of the working force are obtained from the decennial population census and the same, therefore relate to the benchmark years 1951, 1961 and 1971. The year to year estimate derived by following the broad method of interpolation or extrapolation in one or the other form, as the case may be, in each sector. The estimates of the average value of output per worker are obtained for a particular benchmark year on the basis of the information available from the relevant sample surveys conducted specifically for this purpose. The year-to-year estimates are derived by following the broad method of interpolation or extrapolation in one or the other form as the case may be, in each sector. The estimates of the average value of output per worker are obtained for a particular benchmark year on the basis of the information available from the relevant sample surveys conducted specifically for this purpose. The year-to-year estimates at the constant base period prices are obtained by carrying the (1960-61) base year estimates for each sector forward and backward with the help of appropriate physical indicators related to real income in the respective sectors.

The estimates for a given year for each of the different sectors so obtained are then aggregated to arrive at the estimates of the gross domestic product and the net domestic product in the economy as a whole. The estimates of the gross national income product and the net national product are in turn, obtained by adding the net factor income from abroad to the corresponding domestic aggregates.

The following table presents a summary of the braod methodology of national income estimation followed in India:

## An outline of the methodology of National Income estimation in India

| Sector | Nature of available data | Description of the method of estimation |
|---|---|---|
| **Category A**<br>Agriculture, Forests, and Logging, Mining, and Quarrying Registered manufacturing construction.<br>**Category B**<br>Electricity, Railways, Air Transport, Organised Road and Water transport. Communication, banking and Insurance, Real estate, Public | Annual figures of commodity wise output and prices and value of different types of inputs of input - output proportions.<br>The actual figures of all types of factor earnings reported in the annual accounts of different companies or undertaking published on a more or less regular basis. | **Product approach**<br>Aggregation of commodity wise price output multiplications yield gross value of output from which total value of output from which total value of input is detected to arrive at the estimates of value added.<br>**Income approach**<br>The actual figures of employee compensation, |

| | | |
|---|---|---|
| Administration and Defence.<br><br>**Category C**<br><br>Unregistered manufacturing, Gas and Water supply, Unorganised Road and Water transport, Storage, Trade, Hotels and Restaurants, Ownership of Dwellings, Other services. | The estimates of working force derived from the decennial population census data and the estimates of average productivity of labour derived from the data shown by periodical sample surveys. | interest, rent and operating surplus or profits relating to different enterprises / organisations in each sector are aggregated to arrive at the estimates of value added.<br><br>**Income approach:**<br><br>Decennial estimates of working force are interpolated / extrapolated and the periodical estimates of average productivity are carried forward or backward by using certain indicators: The year-to year estimates of workers and their average productivity so derived, are then multiplied to arrive at the estimates of value added. |

## Difficulties in the Estimation of India's National Income

The task of computing national income is a gigantic one and every country, however advanced, faces some difficulties in computing national income. The underdeveloped economies are faced with some special difficulties which arise because of their peculiar nature. In the case of our country, the important problems or difficulties that are faced while calculating national income are:

1. **A large unorganized and non-monetized sector:** While calculating national income, the assumption normally made is that the bulk of the commodities and services produced are exchanged for money. In case of India, however, a considerable portion of output does not come into the market at all being either consumed by the procedures themselves or bartered for other commodities or services. The estimates by the Reserve Bank suggest that in India the non monetized sector is more than one-third of Indian economy. Since transactions in this sector do not come within the purview of money. (There being barter system. Commodities exchanged with commodities) and their valuation for purpose of nationalincome measurement poses grave problems. Moreover, in India, agriculture is mainly done on a subsistence basis and therefore a considerable portion of the output does not come to the market for sale and is consumed by the producers themselves. There is no suitable formula by which the output which is not exchanged for money may be evaluated in terms of money. If this portion is ignored, the national output would be reduced considerably. Hence one big difficulty is to find out the imputed value of non-monetized sector and add this value to the, monetized sector.

2. **Lack of accurate and adequate data:** In the absence of adequate and reliable statisitcs no correct picture of national income can be formed, Statistics in India, particularly for the unorganized sector, are not only inadequate but also very much unreliable. There are no statistics avaialble regarding income, capital investments consumer expenditures, hoarding, customs and excise, domestic employment, cottage industries etc. The main data collecting agency in the village is the village officers. They are semi-literate and untrained in the art of collection of data. Moreover, the collection of data is not their primary function. As a result of it, the statistics collected cannot be very much relied upon. As pointed out by the National Income Committee "the relative dearth of material, both statistical and analytical in the national income field, in India, is part of the vicious circle characterisitcs of an underdeveloped economy, poverty leading to perpetuation of poverty". It is heartening to note that after Independence, continuous efforts are being made by the Government both in respect of improving the reliability of statistics and also the coverage. The Government has set up statistical organisations in every ministry and training is being imparted in the technique of data collection analysis and processing. The Government of India has passed the Collection of Statistics Act, 1953, authorising them to call for information on the various aspects of the Indian economy. Several State Governments have set up statistical Bureau. Statistics of production and output on which national income estimates are based are now available in a much better form in India. However much remains to be done in this respect.

3. **Absence of accounting records:** the problem of measurability is further aggravated by the fact that most producers do not generally maintain any reliable records either of the quantity or of the value of their output. Most of these producers carry on production at family level of run household enterprises on a very small scale. Illiteracy, lack of training and the general absence of the practice of keeping accounts prevents a vast majority of cultivators and small household producers to keep and maintain proper accounts of their respective activity. Not only this, people in our country are not yet alive to the needs of accurate statistics. The result is they are not in a mood to supply information. They do not cooperate with the schemes of the Government in collecting data and are even hostile to them. Thus, peculiar psychology of people is an additional difficulty in the calculation of national Income. Commenting on the non-availability of data the National Income Committee pointed out "an element of guesswork inevitably enters the assessment of output, especially in the large sectors of the economy which are dominated by the small producer or the household".

4. **Regional diversities:** Diversities in different parts of the country are so great that data which are collected in a particular region cannot be used to draw conclusions regarding any other region or the whole of the economy. Our country is a sub-continent and there are great diversities in the climatic conditions, tastes, habits, customs and environments of the people of the different state. Just as statisitcs collected in Japan cannot be used to draw conclusions about Bihar or UP or any other part of the country. The great heterogeneity in different parts of the country makes the collection of statistics necessary for the whole of the country. Even the use of sampling techniques has to be made with great caution. To quote the National Income Committee Report, "Regional diversities in India with its size and varied history are large; and inadequacy of

data cannot easily be overcome by extending data for one part of the country to the rest of the country".

5. **Overlapping of economic functions:** A major difficulty in the calculation of national income in our country is the lack of clear-cut line of demarcation between different occupations. Most of the people in our country particularly in the rural areas follow more than one occupation and it becomes difficult to allocate their income to deferent occupations. This difficulty arises because of the fact that agriculture in India is seasonal and as such during the off season the agriculturist engages himself in a number of pursuits such as barber, carpenter, labourer or else, he often migrates to the town and works as a domestic servant, as a rickshaw puller, hawker or an industrial labourer. As a result it is not possible to arrive at a suitable occupational classification and the computation of national income by industrial origin becomes very difficult.

## Comparison of National Income Aggregates over time

The National income is measured in terms of the usual measuring rod of money. Current money prices are used as appropriate "weights" to aggregate the annual output of diverse products produced in the economy. Measurement of national income, therefore, depends upon two types of factors (1) quantities of different products actually produced during the given year and (2) the corresponding set of money prices used for converting diverse physical quantities into standardised values for the purpose of aggregation. If the latter also relates to the same year as the former. The resulting aggregate is called national product at current prices, or Money National Product (money NNP)

If we have the figures of money NNP for two years, say 1980 and 1990, a direct comparison of two figures will show the direction and magnitude of change in the aggregate flow of factor incomes originating in the economy between 1980 and 1990. However, it will not give us any indication regarding the magnitude and in many cases even the direction of the underlying change in the volume of physical output produced in the economy between the two years under consideration. The reason is simple. Both the type of factors mentioned above that determine money NNP, viz., physical output and money prices, are likely to have undergone a change between 1980 and 1990. To measure the change in the physical output, we should eliminate the effect of changes in the price level.

The measure that is devised for the purpose of comparing the volume of physical output produced during different years is generally known as Real National Product (Real NNP). This is derived by "deflating " the money NNP with an "index number prices". Deflation is the produce by which the effect of variations in the measuring rod of money prices is "eliminated". Index number of prices or "price index" is a measure of the average degree of price changes between the two years under consideration. It represents a weighted average of the relative changes in the prices of various products, the weights being assigned in accordance with the measured economic importance of the corresponding products. The year is relation to which the relative price changes are computed for another given year is referred to as the 'base year'. The price index for the base year is taken as 100. Thus for instance, if the price index for 1990 with 1980 as the base year (i.e. the price index for 1980 being 100) is 175, it implies that on an average, prices increased by 75

percent in 1990 as compared to 1980. Viewed from another angle the price index shows the extent by which money NNP would have increased had the real output of goods and services remained constant during the period under consideration.

If we have the figures for money NNP for different years and if we are also given index numbers of prices with a given base year, we can readily derive the so called real NNP by using the following simple formula:

$$\text{Real NNP} = \frac{\text{Money NNP}}{\text{Price Index}} \times 100$$

The derivation of real NNP can be illustrated with the help or a simple numerical exercise as shown in the following Table. The interpretation of the figure Rs.20,000 crores, show in the last column of the Table as Real NNP for the year 1990-91, is quite simple. In represents the value of final goods and service produced in 1990-91 at the prices prevailing in 1980-81. In other words, this calculation shows that if prices on an average had remained unchanged between 1980-81 and 1990-91, national income in 1990-91 would have been only Rs.20,000 crores and not Rs.35,000 crores as indicated by the figure of money NNP for 1990-91. Thus, real NNP can be regarded as an approximate measure of the extent of change in the volume of real output produced in the economy.

**Deflation of money NNP to compute Real NNP**

| Year | Money NNP (Rupees in crores at current prices) | Index numbers of prices | Real NNP (Rupees crores at constant 1980-81 prices |
|---|---|---|---|
| (1) | (2) | (3) | $(4) = \frac{(2)}{(3)} \times 100$ |
| 1980-81 | 14,000 | 100 | $\frac{14,000}{100} \times 100 = 14,000$ |
| 1981 | 70 | 110 | $\frac{35,000}{175} \times 100 = 120,000$ |

## International Comparison of National Income Aggregates

The main purpose behind international comparisons of national income is to get a clear idea of the extent to which aggregate output and general living standards actually differ from one country to another. In practice, however, comparisons of national income aggregates between different countries are more complex than comparison of national income over the years for a given country. Many difficulties, both conceptual and statistical arise when we try to compare national incomes of different countries.

Firstly, the form, content and scope of economic activity itself may vary considerably from one country to another. For each country, the institutional framework, social customs, cultural factors and the level of economic development determine the nature and extent of non-marketed activity will show significant inter-country variations thereby making the task of international comparisons of national income quite difficult.

We can illustrate this argument by giving a specific example. It is quite possible for instance, to observe that in a developed country many services such as motor driving, house painting, furniture repairing, tailoring, etc., are rendered by the members of a household in the form of non-marketed personal services, eventhough in the same economy such services also tend to be performed on a commercial basis. Now, this may not be the case in an underdeveloped country with a semi-feudal social structure and the custom of having paid domestic servants, especially in households belonging to higher income groups. It is obvious that the practice of excluding non-marketed activity and personal services from national income aggregates will have a definite impact on the direct comparison of national incomes of the two countries in this case.

Secondly, in most cases the imputed vale of non-marketed output forms a significant part of the national income. Eventually, the methods of imputation are likely to differ from country to country. Similarly, the treatment of general government output may vary from one country to another. Such differences, wherever they exist, will obviously make the reported national income aggregates non-comparable.

Thirdly, different countries generally have different currencies and this gives rise to the problem of converting the money values of their national income aggregates expressed in terms of their respective currencies to a common monetary basis for making direct comparisons possible. It is customary to use the official exchange rates between different currencies for this purpose. However, the use of exchange rates creates serious difficulties in introductory comparisons because the official exchange rate does not measure in any absolute sense the true exchange value between the two currencies involved.

The exchange rate, at best, is capable of reflecting the relative valuations in terms of different currencies of only those commodities that enter into international trade after making due allowance for the influence of international capital movements. Thus, for instance, the official exchange value at par say, of one US dollar being equal to 30 Indian rupees, will imply that an Indian household with a monthly income of Rs.3000 is on par in terms of its general living standards, with an American household having a monthly income of 100 dollars, which in practice is not likely to be true.

To avoid the problems associated with the use of official exchange rates it is customary to take recourse to the concept of purchasing power parity. This concept serves the purpose of defining a suitable ration of exchange between different currencies on the basis of the general purchasing power of each currency which is expressed in terms of the amount of goods and services it can buy in a domestic market. From a purely theoretical view point, the purchasing power parity can be regarded as a fairly satisfactory device for converting national incomes expressed in different currencies to a common currency base. However, from a strictly statistical

angle this concept is quite elusive because in actual practice detailed commodity by commodity comparisons of the purchasing power of two different currencies in their respective domestic markets are extremely difficult.

Another alternative is to choose a single comprehensive set of prices and value the output of different countries in terms of this. However, this alternative is also quite difficult to operate in actual cases because the commodities produced in different countries are too diverse and heterogeneous to be treated as even broadly similar. In many cases we may find that a number of products produced and consumed in fairly large quantities in one country are just not produced in another country. Moreover, even in the case of those goods and services that can be taken as apparently identical, we may find significant quality differences between the products actually produced in different countries.

In addition to all this, we may also find that the relative prices vary considerably from one country to another. Relative price differentials will pose the problem of choosing the appropriate price basis. This choice represents the dilemma typical of the index number problem. It is an important problem because it invariably affects the overall comparison of national income aggregates to a significant level. Thus, neither straight evaluation at a single set of prices nor any derivation of general purchasing power parity based on straight price comparisons seems to be free from serious difficulties and drawbacks.

It is obvious from the above discussion that to compare the national income aggregates of different countries is a fairly difficult task. One must, therefore, be quite cautious in attempting such comparisons and interpreting their results.

## Recent improvements in national income estimates

It is heartening to note that serious efforts are being made to remove these difficulties. Periodically a conference on national income is held to consider the various methodological improvements in the calculation of national income in India. The advice of the experts from India and abroad is being sought on various problems. In this way a great headway is being made in making these estimates about India more realistic and accurate. Amongst the important steps taken up by the Government to improve the national statistics, special mention may be made of the following:

1. Setting up of the National Income Committee in 1949. This was a landmark in the history of National Income statistics of our country.

2. The creation of the National Income Unit with the Central Statistical Organisation with the responsibility of compiling national income data and bringing about the continuous improvements in them.

3. With a view fill in the statistical gaps, a number of new studies have been made both by CSO and other institutions. Important among these are: Estimates of gross capital information in India for 1948-49 to 1960-61. Preparation of the 'Input - Output'. Table by the Indian Statistical Institute, Monograph on Savings in India 1948-49 to 1957-58 by NCAER, Studies on Saving and Finances of Public and Private Limited companies, Estimates of net tangible wealth of India. (1949-50 to 1960-61) by the Reserve Bank of India.

4. In order to reduce the time lag of ten to twelve months between the expiry of the financial year and the publication of the white paper on national income, the CSO had started issuing 'quick estimates' from year 1958-59 onwards. The latest 'quick estimates' relating to the year 1991-92 have been recently released.

5. The CSO has set up a Working Group on State Incomes and efforts in this direction have improved the coverage, regularity and accuracy of the work done by the state statistical Bureau.

6. From 1961 onwards, budget documents of public authorities have been used to provide not only the 'actuals' but also the revised estimates, so that up-to-date information is provided by the National Income Tables. Similarly agricultural crops are being valued by actual state wise price observations rather than by projecting year's weights all India average prices on the basis of the Economic Adviser's wholesale Price Index Number of Agricultural Commodities.

## Recent Estimates - Trends

The following Table presents the estimates of India's national income and its related aggregates. The analysis of these estimates unfolds the following facts.

The GDP* at factor cost at current price rose from Rs.1,22,226 crores in 1981-82 to Rs.2,92,707 crores in 1986-87. Thus, GDP at current prices rose approximately by 2.5 times over 1980-81 to 1986-87. Over the same period GDP at constant prices of 1980-81 rose from Rs.1,22,226 crores to Rs.1,64,441 crores ie., by 1.25 times. In other words, GDP in money terms rose by 2.5 times, whereas in real terms by 1.25 times over 1980-81 to 1986-87.

Similarly, NDP at factor cost at current prices recorded an increase from Rs.1,10,139 crores to Rs.2,30,464 crores over 1980-81 to 1986-87. Over the same period NDP at constant price, of 1980-81 increased from Rs.1,10,139 crores to Rs.1,44,963 crores. Thus, NDP in money terms rose approximately by 2 times whereas in real terms it increased by 1.25 times approximately.

The GNP at factor cost at current prices increased approximately by 2 times over 1980-81 to 1986-87. Over the same period GNP at constant prices of 1980-81 increased approximately by 1.25 times.

## Table
## National Income aggregates at 1986-87 prices and at 1980-81 prices

(Rs. Crores)

| | 1980-81 | 1981-82 | 1982-83 | 1983-84 | 1984-85 | 1985-86 | 1986-87 |
|---|---|---|---|---|---|---|---|
| 1. GDP at Factor cost | | | | | | | |
| a) Current prices | 1,22,226 | 1,59,420, | 1,77,588 | 2,07,272 | 2,29,542 | 1,61,733 | 2,92,787 |
| b) 1980-81 prices | 1,22,226 | 1,29,776 | 1,33,830 | 1,44,817 | 1,50,542 | 1,58,176 | 1,64,441 |
| 2. Consumption of Fixed Capital | | | | | | | |
| a) Current prices | 12,087 | 14,459 | 16,886 | 19,332 | 22,257 | 26,440 | 0,120 |
| b) 1980-81 prices | 12,087 | 12,788 | 13,595 | 14,922 | 17,036 | 18,482 | 19,478 |
| 3. NDP at factor cost | | | | | | | |
| a) Current prices | 1,10,139 | 1,21,417 | 1,41,965 | 1,67,084 | 1,84,47 | 52,06,865 | 2,30,464 |
| b) 1980-81 prices | 1,10,139 | 1,16,988 | 1,20,235 | 1,29,895 | 1,33,506 | 1,39,69 | 41,44,963 |
| 4. GNP at factor cost | | | | | | | |
| a) Current prices | 1,22,571 | 1,42,916 | 1,58,217 | 1,84,871 | 2,06,357 | 2,32,634 | 2,58,875 |
| b) 1980-81 price | 1,22,571 | 1,29,815 | 1,33,214 | 1,43,365 | 1,49,813 | 1,56,975 | 1,62,997 |
| 5. NNP at factor cost | | | | | | | |
| a) Current prices | 1,10,484, | 1,28,457 | 1,41,331 | 1,65,642 | 1,84,273 | 2,06,419 | 2,29,010 |
| b) 1980-81 prices | 1,10,484 | 1,17,027 | 1,19,619 | 1,29,344 | 1,33,972 | 1,40,647 | 1,45,675 |
| 6. Per capita NNP (Rs) | | | | | | | |
| a) Current prices | 1,627 | 1,851 | 1,993 | 2,287 | 2,403 | 2,734 | 2,974 |
| b) 1980-81 prices | 1,627 | 1,686 | 1,687 | 1,786 | 1,862 | 1,862 | 1,891 |

Source: Compiled from RBI's Report on Currency and Finance 1987-88, Vol. II. Statements 3-4, pp. 6-7; Government of India's Economic Survey, 1988-89 pp. 5-3 and 5-6

The NNP at factor cost (National Income) at 1986-87 prices ie. money national income increased by 2.07 times over 1980-81 to 1986-87, whereas NNP at constant prices (1980-81), ie., real national income increased by 1.31 times over the same period.

The per capita income at current prices, ie. money percapita income increased by 1.82 times over 1986-87, whereas the percapita income at constant prices (1980-81) ie. real percapita income increased by 1.16 times over the same period.

There was a very perceptile improvement during 1980-81 to 1990-91. Net National product showed a growth rate of 5.4 percent per annum and the per capita NNP at (1980-81 prices) improved on an average by 3.1 per cent per annum. During 1990-91 to 1997-98, the average

*GDP (Gross Domestic Product) is the aggregate money value of goods and services produced by people living within the geographical boundaries of a nation.

annual rate of growth, of net national product at (1980-81 prices) was just of the order of 5.5 per cent per annum and that of NNP per capita was 3.6 per cent. The sharp spurt in growth is attributed to high growth in agriculture, manufacturing, electricity, construction trade and hotels, transport other than railway and communication. During 1993-94 and 1997-98, the growth rate of NNP was 7.0 per cent on the average and of per capita NNP at 5.1 per cent.

**Table**

**Trends in Growth Rates (Percent)**

| | GNP at factor cost | | NNP at factor cost | | Percapita NNP | |
|---|---|---|---|---|---|---|
| Year Prices | At current Prices | At 1980-81 Prices | At current Prices | At 1980-81 Prices | At current Prices | At 1980-81 Prices |
| 1981-82 | 16.6 | 5.9 | 16.3 | 5.9 | 13.8 | 3.6 |
| 1982-83 | 10.6 | 2.6 | 10.0 | 2.2 | 7.7 | 0.1 |
| 1983-84 | 16.8 | 8.0 | 17.2 | 8.1 | 14.8 | 5.9 |
| 1984-85 | 11.6 | 3.9 | 11.2 | 3.6 | 9.8 | 1.5 |
| 1985-86 | 12.7 | 5.1 | 12.0 | 5.0 | 9.6 | 2.8 |
| 1986-87 | 11.3 | 3.8 | 10.9 | 3.6 | 8.8 | 3.6 |
| 1987-88 | 12.6 | 3.6 | 12.6 | 3.4 | 10.4 | 3.4 |

Source: Government of India, Economic Survey, 1988-89 p. 55

The trends in national income and percapita income can also be analysed in terms of growth rates as appended in table: Some of the important findings are as follows:

1. During 1980-81 to 1987-88, the year 1983-84 witnessed the highest growth rates of 16.8 percent and 17.2 percent in GNP and NNP at current prices respectively. In real terms (at 1980-81 prices) also, the year 1983-84 recorded highest growth rates of 8 percent and 8.1 percent in GNP and NNP respectively. These highest growth rates were largely due to an accelerated increase in agricultural production. The agricultural production, during 1983-84, increased at the rate of 13.7 percent as compared to a negative growth at 3.3 percent during the preceding year. The industrial production also recorded a rise by 6.7 percent during 1983-84 as compared to marginal rise at 3.25 percent during the preceding year.

2. During 1980-81 to 1987-88 the year 1982-83 could be identified as the year of deceleration in GNP and NNP both at current and constant prices (1980-81). During 1982-83, the growth rates of GNP and NNP at current prices were of the order of 10.7 percent and 10.0 percent respectively. In real terms (1980-81 prices) also, the growth rates of GNP and NNP decelerated to 2.6 percent and 2.2 percent respectively as compared to 5.9 percent growth rates of both GNP and NNP in the preceding year, 1980-82. The deceleration in the growth rates of GNP and NNP could be explained in terms of the poor performance of agricultural, industrial and power sectors.

The agricultural sector experienced a negative growth of 3.3 percent as compared to 5.6 percent growth in agricultural production in the preceding year, 1981-82. Similarly, during 1982-83 industrial production increased by 3.2 percent as compared to a rise by 9.3 percent in the preceding year. The power generation recorded a growth of 6.7 percent as compared to the growth rate of 10.2 percent in the preceding year 1981-82.

3. During 1981-82 to 1987-88, the real percapita NNP has increased at annual average rate of 3.4 percent despite annual growth of population of 2.2 percent during 1981-87. The appreciable growth in real per capita income has become possible to 5.3 percent annual average growth in real NNP during 1981-82 to 1987-88. However, it is important to note that the per capita income has shown a rather zig zag movement dispite steady growth in population.

4. During 1981-82 to 1987-88, annual average growth in real GNP and real NNP were of the order 5.5 percent and 5.3 percent respectively. Thus, the annual average growth rates of real GNP and real NNP, during 1981-82 to 1987-88, were commendable as these were in consonance with the targeted growth rates of the Sixth and Seventh Five Year Plans.

The National Accounts Statistics in 1999 estimated average growth of GDP of the order of 7.7 per cent. But this growth rate declined subsequently to 5.7 per annum during 1993-94 - 1998-99. Therefore during this period the growth rate per capita NNP was 4.1 per cent.

During 1993-94 to 1997-98, the share of the primary sector in GDP has gone down from 30.3 percent in 1993-94 to 26.4 per cent in 1997-98 at constant prices. The share of the secondary sector and tertiary sector has increased.

The per capita private final consumption expenditure which was Rs. 6,366 in 1993-94 has increased to Rs. 7490 in 1997-98.

There is a structural change for the better in the composition of national income. It is due to the process of economic growth initiated during the plans. The growth process involved a rapid expansion of manufacturing in the organised sector. However, agricultural sector did not indicate a fast rate of growth 1950-51 to 1990-91, the average rate of growth of agriculture was 2.6 per cent per annum, whereas the growth rate of manufacturing sector stood at 5.6 per cent per annum during the same period. The rate of growth of transport and communications was of the order of 5.4 per cent, banking, insurance and real estate stood at 4.4 per cent.

During 1990-91 to 1996-97, during the period of new economic reform agricultural growth rate was to the extent of 2.8 percent and of the service sector went up by 7.4 per cent.

The structural change in the composition of India's national product indicates a higher share of industry and services and relatively a lower share for agriculture. The disparity in per capita incomes between developed and developing countries is a reflection of the disparity in the structure of the nature. The developed economies are largely industrial in nature while the developing economies are largely agricultural.

Since the onset of the plans and mainly since the post reform period of 1990-91 onwards Indian economy is passing through a process of transition from an agricultural economy to an industrial one. In this process, a structural change in the composition of national income has taken place.

The changing structure of national income needs to be strengthened by stepping up the programme of industrialisation. At the same time accelerating the growth process in agriculture, industrialisation of the economy with emphasis on agro-based industries and industries supplying inputs to agriculture is a must.

## Social Accounts and the Measurement of National Income

Business enterprises keep record of their transactions in the form of accounts. Amongst other things these accounts provide information upon which the business enterprises can base policy decisions. They provide a continuous commentary on the affairs of business concerns and are essential to management in controlling efficiency, in ensuring financial stability and satisfactory rate of profit, and in facilitating balanced growth.

Just as there are many reasons which make it necessary for individual businesses to keep records of their transactions, so it is most important to have records of the transactions which take place in the national economy as a whole. These records are called social accounts. These accounts provide information about the structure and functioning of the economy. If appropriate classifications are made information can be derived about the annual income of the nation, how it is produced, distributed and spent, how the wealth of the nation is being built up and so on. Such information provides a basis for national economic stability and prosperity and to ensure an efficient distribution of economic resources and balanced growth of the economy. Not only this, empirical investigations of the working of an economy depends on the availability of data about aggregates of transactions of the kind recorded in social accounts.

It was after the publication of Lord Keynes' The General Theory of Employment, Interest and Money, in 1936 that social accounting received attention although its recent development is associated with the name of Dr. J.K.N. Stone and Prof. Meade. According to Edeg and Peacock: "Social Accounting, embracing national income accounting is concerned with the statistical classification of the activities of human beings and human institutions of economic activity and also the application of the information thus assembled to the investigation of the operation of the economic system".

Social accounts are so drawn as to bring out the distribution between i) forms of economic activity, ie. production or consumption or accumulation of wealth, ii) sectors or institutional subdivision of the economy, ie. private or public sector and iii) types of transactions, ie. sales, purchases, gifts, taxes etc.

## Uses of the Social Accounting Framework

Economists have always been interested in aggregates such as national income but the systematic treatment of aggregate transaction in a social accounting framework dates only from

about 1940. The following are the uses of the social accounting system:

1. It enables the structure of economic transaction to be set out in a consistent way and make clear the dependence of the definition of any given aggregates on the particular system choses. It helps to elucidate the relations between associated concepts, ie. the distinction between gross national product at market prices and net national product at factor cost. It readily reveals the effects of any change in the treatment of particular items on the various aggregates of transaction.

2. Social accounts facilitate the actual estimation of the aggregates of transaction. They indicate alternative routes to the estimation of a given concept, ie. the estimation of gross national product by the production, expenditure and income received method. Any discrepancies and errors are revealed and a basis is provided for making adjustments of discrepancies.

3. Social accounts provide a framework for the classification of transactions and hence suggest the form in which data should be collected. Also movements in gross national product valued at constant prices and expressed per head of the population give an indication of movements in the standard of living. Similarly, movements in gross national product, valued at constant prices and expressed per head of the working population, give an indication of movements in the level of productivity.

4. The social accounts give a picture expost of the outcome of economic activity. Expost measurement related to events which have actually taken place. They are a record of what has happened. They can also be used as framework for drawing up an ex-anteforecast of the likely outcome of the economy in the future. Ex-ante measurements relate to expectations, intentions, plans, forecasts etc.

5. The social accounts method gives a better understanding of the state of nation's economy, its functioning, its break up and how the break up values stand in relation to each other. The method of accounts channels of the policy maker away from the immediate pressures of the specific problems to a consideration of the whole economy of the close inter relation among its parts and of the links between the present and future.

## Social Accounting in India

In India, we do not have adequate statistics to have a meaningful system of social accounts. The National Income Committee, as early as in 1948, pointed out that "social accounting as a method of estimating the national icnome and of studying the trends in the various sectors of India's economy is full of difficulties in view of the lack of statistical data available on certain important aspects of the Indian economy". The Committee found that data relating to depreciation, capital information, investment, savings and consumer expenditure, etc., were not satisfactory. The Committee worked with the poor data at their command and presented a simple framework of social accounts for India. A brief description of the system is given below:

The National Income Committee adopted a classification of the Indian economy into the following three basic sectors.

i) Enterprise,

ii) Household and

iii) Government

Each of the three sectors were further subdivided into four heads as follows:

a) Production

b) Consumption

c) Addition to wealth and

d) Social Accounts

The entire framework of social accounts, therefore, consists of twleve accounts in all comprising three major sectors of the economy, classified further into four sub classes. The method of presentation in a tabular form is as follows:

## Framework for Social Accounts

| Sectors of Accounting | Economic Activities | | |
|---|---|---|---|
| | Enterprises | Household | Government |
| Production | | | |
| Consumption | | | |
| Addition to Wealth | | | |
| External Accounts | | | |

**Account I : Domestic Product**

| | |
|---|---|
| 1.1 Gross domestic product at factor cost (2.9) | |
| 1.2 Indirect taxes (5.7) | 1.6 Gross domestic fixed capita formation (3.1) |
| 1.3 Less subsidies (5.2) | 1.7 Increase in stocks (3.2) |
| 1.4 Private consumption expenditure (4.1) | 1.8 Exports of goods and services (6.1) |
| 1.5 Central Government consumption expenditure (5.1) | Expenditure on gross domestic product and imports |
| | 1.9 Less imports of goods and services. (6.3) |
| Gross domestic product market prices at | Expenditure on gross domestic product |

## Account 2: National Income

| | |
|---|---|
| 2.1 Compensation employees (4.5) | 2.9 Gross domestic product at factor cost (3.1) |
| 2.2 Income from farms, professions and other unincorporated enterprises (4.6 | 2.10 Net factor income payment from the rest of the world (6.2) |
| 2.3 Income from property (4.7) | 2.11 Less provisions for the consumption of fixed capital (3.3) |
| 2.4 Savings of corporations (3.4) | |
| 2.5 Direct taxes on corporations (5.8 | |
| 2.6 General government income from property and entrepreneurship (5.5) | |
| 2.7 Less interest on the public debt (5.6) | |
| Less interest on consumer, debt (4.8) | Net national product at factor cost |
| National Income | |

## Account 3: Domestic Capital Formation

| | |
|---|---|
| 3.1 Gross domestic fixed capital formation (1.6) | 3.3 Provisions for the consumption of fixed capital (2.11) |
| 3.2 Increase in stocks (1.7) | 3.4 Savings of corporations (2.4) |
| | 3.5 Net capital transfers from household and private non-profit institutions (4.11) |
| | 3.6 Net capital transfers from general government (3.11) |
| | 3.7 Net International transfers received by corporation (6.6) |
| | 3.8 Net borrowings of corporation (4.14+5.15+6.9) |
| Gross domestic capital formation | Finances of gross domestic capital formation |

## Account 4: Households and Private Non-profit Institutions:

| | |
|---|---|
| 4.1 Consumption expenditure (1.4) | 4.5 Compensation of employee (2.1) |
| 4.2 Direct taxes (5.9) | 4.6 Income from farms professions and other unincorporated enterprises (2.2) |
| 4.3 Other current transfers to general government (5.10) | 4.7 Income from property (2.3) |

| | |
|---|---|
| 4.4 Savings (4.12) | 4.8 Less interest on consumers debt (2.8) |
| | 4.9 Current transfers from general government (5.3) |
| Disposal of income | Income of household and private non-profit institutions. |

**Capital Reconciliation Account**

| | |
|---|---|
| 4.10 Net capital transfers to general government (5.13) | 4.12 Savings (4.4) |
| 4.11 Net capital transfers to domestic capital formation (3.5) | 4.13 Net international transfers received (6.7) |
| | 4.14 Net borrowing (3.8 + 5.15+6.9) |
| Dibursements | Receipts |

**Account 5: General Government**

| | |
|---|---|
| 5.1 Consumption expenditure 91.5) | 5.5 Income from property and entrepreneurship (2.6) |
| 5.2 Subsidies (1.3) | 5.6 Less interest on the public debt (2.7) |
| 5.3 Current transfer to households (4.9) | 5.7 Indirect taxes (1.2) |
| 5.4 Savings (5.12) | 5.8 Direct taxes on corporations (2.4) |
| | 5.9 Direct taxes on households (4.2) |
| | 5.10 Other current transfers from households (4.3) |
| Disposal of revenue | Current revenue |

**Capital Reconciliation Account**

| | |
|---|---|
| 5.11 Net capital transfers to domestic capital formation (3.6) | 5.12 Savings (5.4) |
| | 5.13 Net capital transfers from households (4.10) |
| | 5.14 Net international transfers received (6.8) |
| | 5.15 Net borrowings (3.8+4.14+6.9) |
| Disbursements | Receipts |

**Account 6: External Transactions (Rest of the World Accounts) Current Account**

| | |
|---|---|
| 6.1 Exports of goods and services (1.8) | 6.3 Imports of goods and services (1.9) |
| 6.2 Net factor income payments to the nation (2.10) | 6.4 Surplus of the nation on current account (6.5) |
| Current receipts from abroad | Disposal of current receipts from abroad |
| Capital Reconciliation Account | |
| 6.5 Surplus of the nation on current account (6.4) | 6.9 Net lending to the rest of the world (3.8+4.14+5.5) |
| 6.6 Net international transfers to corporations (3.7) | |
| 6.7 Net international transfers to households (4.13) | |
| 6.8 Net international transfers to general government (5.14) | |
| Receipts | Disbursements |

Commenting on the relationships among the entries on the accounts, the U.N. Series F.No.2 Rev.1 gives the following derivations.

i) National Income - Net-domestic product at factor cost plus net factor income from the rest of the world.

ii) National Income - Consumption expenditure plus net domestic capital formation plus net export plus net factor income from the rest of the world less indirect taxes net of subsidies.

iii) National Income - Consumption expenditure plus net domestic capital formation plus surplus of the nation on current account less net current transfers from the rest of the world less indirect taxes net of subsidies.

In order to follow this set of Standard Accounts, we will have to do a lot with regard to the availability of reliable data on many of the aspects on which no data or very little data at present are available.

## Model Questions

1. Explain the process of calculating national income in India and the difficulties in calculating it.
2. Describe the social accounting method and its uses.

## LESSON -14

# INDIAN OFFICIAL STATISTICS
# NATURE AND STRUCTURE OF THE OFFICIAL STATISTICAL ORGANISATION

The term official statistics is used to describe the vast quantity of primary and secondary data which the state collects, compiles and publishes in official publications either for information of the public or for use in related Government departments.

The nature and structure of the official statistical organisation can best be understood in the background of the administrative system, including the Constitution of India. India, being a federation of states, there is a dichotomy of responsibility for Government between the Union or Central Government on the one hand and the Governments of states on the other. Thus the present statistical system in India is a decentralised one. Under the Indian Constitution, the responsibility is divided on the basis of a three fold classification of all subjects, viz, Union List, State List and Concurrent List, the last category representing the areas where both the Union and State Governments could operate. Some of the items under the Union List are Defence. Railways, Posts and Telegraphs, Currency and Foreign Exchange, Trade and Commerce and Foreign affairs Census, Income-tax, Customs and Excise duties, etc. The Central Government is responsible for the collection of statistics for the items in the Union List. Some of the items under the State List are Public Health, Agriculture, Livestock, Irrigation, Forest and Fisheries etc. The State Governments are responsible for the collection of statistics for these items. The Concurrent List includes Vital Statistics, Industry, Social Insurance, labour welfare, Relief and Rehabilitation, Price control, etc. In these fields, both the Central and state Governments can operate. At the central as also the state levels, there is a further division of responsibility by subject of groups of subjects among the different Central Ministries, and State Government Departments' responsibility. Thus the present statistical system in India is a decentralised one, in which the authority and responsibility for collection of statistics is divided between the Union or Central Government and State Governments on a subject -wise basis, with a further division, subject-wise again, of both the central and state levels.

However, appropriate to the federal set-up, the Central Government acts by understanding the co-ordinating agency for presentation of statistics on an all-India basis, even in fields where the states have the primary authority and responsibility for collection of statistics. At the Centre a Central Statistical Organization (CSO) exists for co-ordination at the national level of the activities of the State and Central statistical agencies. At the State level, the State Statistical Bureau attached to different departments in different state governments are charged with the responsibility of co-ordination of all statistics at state level and keeping liaison with the CSO for purpose of co-ordination at All India level. The CSO is the technical wing of the department of statistics located in the Cabinet Secretariat.

At the Centre, most of the ministries have either a full fledged statistical department, a division or a section depending upon their needs and upon the stage of development of statistics in the

relevant fields. Many of these had been set up long ago and had been fitted into their appropriate places in the statistical system organised since Independence. At the state level, the State Government Department with responsibility for different subjects contain statistical units big or small concerned with collection and compilation of statistics relating to the particular field. But as mentioned earlier, even for state subjects, responsibility for co-ordination and compilation of all-India figures rest with the concerned Central agencies.

## Statistical Organisation at the Centre

At the Centre, most of the Ministries collect or use statistics in some manner or other and each have its own statistical unit, (some have more than one) which is responsible for collection and compilation of statistics relating to the subject of the Ministry. These statistical units are of different sizes, and in varying stages of development and are charged with distinctive functions. The official statistics are collected by anyone of the following ways:

### (1) Organizations processing data that are obtained as a by-product of administration

Some statistical units located in the administrative departments are engaged in processing of data which are purely by -products of administration. Examples of such agencies are Offices of Income-tax, Central Board of Revenue, Railways, Posts and Telegraphs and the Directorate General of Supplies and Disposals. The state is running many business units also. In the natural course of administration, data are collected for the purpose of efficient administration. These units simply analyse and interpret them and provide the information to the related departments.

### (2) Statistical Organisation attached to central agencies

There are some statistical units in organizations set up for control of production and distribution products in short supply and these maintain statistics which are valuable to government organizations and public.

Examples of these are : Textile Commissioner's Office, Central Excise Commissioner's Office, Iron and Steel Controller etc. These organisations are not required to collect data, but they are provided with the data that were actually collected by other agencies for the purpose of maintaining control over production etc.

### (3) Research Organizations ( Statistical)

There are a few statistical units at the centre level, associated with some major branches of operation that have been assigned the duty of statistical research in their respective branches. Some of them are :

a) Research Department of the Reserve Bank of India.

b) Statistical Division of the Indian Council of Agricultural Research.

c) Research Section of the Indian Standard Institute.

(4) Statistical Organizations Specially Created For The Collection of Data

There are also various organizations established by the Government, specially for the function of collecting, compiling and co-ordinating statistical data. These are :

i) National Census Organization and the Office of the Registrar General of India (Ministry of Home Affairs)

ii) Directorate General of Commercial Intelligence and Statistics (Ministry of Foreign Trade)

iii) Labour Bureau (Ministry of Labour, Employment and Rehabilitation)

iv) Directorate Economics and Statistics (Ministry of Agriculture and Irrigation)

v) The Army Statistical Organisation (Ministry of Defence)

vi) Directorate of National Sample Survey (Cabinet Secretariat)

vii) Central Statistical Organisation (Cabinet Secretariat)

A brief description of thee organisations and units is given below :

## i) National Census Organisation And Office of the Registrar General and Census Commissioner of India (Ministry of Home Affairs)

Beginning with 1881, the census has since been taken once in every ten years. The census operation is of immense magnitude, carried out with the help of about 8 million honorary enumerators including Governments employees, employees of local bodies and teachers. Plans have been worked out by the Registrar General and Census Commissioner for evolving and integrated system of demographic and vital statistics on a continuing basis and these are under implementation.

Besides bringing out the Census Reports the Office brings out.

i) Vital Statistics of India (Annual) and

ii) Indian Population (Biennial)

## ii) Directorate General of Commercial Intelligence and Statistics (Ministry of Foreign Trade)

Set up in 1895, this Organisation was called, till the Second World War, the General Statistical office responsible for collection, consolidation and publication of important statistical series. With the formation of statistical units in the different Ministries during the war, many of the former functions of this office were transferred to the appropriate Ministries and it was left with only commercial intelligence and trade statistics. Now its main functions are :

1. Collection and furnishing commercial information required by the Government and trade,

2. Mediation in commercial disputes between Indian and foreign businessmen with a view to bringing about amicable settlements.

3. Granting trade introductions,

4. Maintaining a register of firms in India,

5. Maintaining a commercial library and reading room for the use of the public.
6. Assisting commercial concerns with a view to stimulating foreign trade of India particularly the export of Indian produce and manufactures.
7. Dissemination of commercial information received from the Government trade representatives abroad;
8. Assisting persons engaged in trade and industry;
9. Compiling and publishing statistics or trade, shipping etc., in the publications issued by the Department.

This office brings out Indian Trade Journal, a weekly.

### iii) Labour Bureau (Ministry of Labour)

It was set up in 1946 with the function of collection, publication and dissemination of statistics on various aspects of labour including construction and maintenance of consumer price index number. It is also charged with conducting research into specific problems with a view to furnishing data required for the formulation of labour policy and labour legislation. The Bureau also conducts budget enquiries for industrial workers and compiles consumer price index numbers for working class.

### iv) Directorate of Economics and Statistics (Ministry of Agriculture and Irrigation)

It was set up in 1947 with the responsibility for compiling and publishing agricultural statistics on an All-India basis. The data covers the fields of agriculture, fisheries, live-stock and forests, and collected by the state Governments partly through administrative channels and partly through specially designed random sample surveys. It also collects agro-economic data and advises the Ministry on matters of agro-economic policy.

### v) The Statistical Organisation (Ministry of Defence)

It is established under the Ministry of Defence and processes all important data relating to personnel. It has one of the largest mechanical tabulation equipment for tabulation of data. There is also a research unit connected with application of sampling methods of surveys etc.

### vi) The Department of Statistics (Ministry of Planning)

The Department of Statistics was formed in the Cabinet Secretariat in 1961. The C.S.O became a part of it along with the NSS Directorate in 1966. Later a computer centre was set up to cater to the data processing needs of the Government departments located in and around Delhi. The Department of Statistics is closely associated in its working with the Indian Statistical Institute, Calcutta. Since 1973 the Department was brought under the Ministry of Planning.

### a) Directorate of National Sample Survey (NSS)

Another important agency at the Centre is the Directorate of NSS. The difficulties in the way of expeditious collection of data on a comprehensive basis, limitations of cost and manpower, brought to the fore the sampling approach. The multi-purpose NSS directorate was set up in 1950

in the Ministry of Finance for collecting data on all aspects of the national economy on a continuing basis and required by the National Committee, Planning Commission and various Ministries of the Government of India. The Directorate of the NSS was transferred to the Cabinet Secretariat in 1957.

The NSS, Directorate has been conducting multi-purpose socio-economic sample survey in the form of rounds and the NSS Reports are regularly (round-wise) published by it. The programme for each round is decided by the Department of Statistics in collaboration with the concerned Ministries and State Governments. The responsibility for technical guidance and tabulation of data rests with the Indian Statistical Institute.

There are three main functions of the Directorate of NSS. They are ;

i) To conduct socio-economic surveys with an All-India coverage by the method of random sampling for collecting data on socio-economic conditions of the population, price, and wages, production in small-scale household enterprises, consumption and agriculture. The data are collected on a continuing basis with a view to filling gaps in the statistical information required for national income estimation, planning and policy-formation and administrative decisions by the different Ministers of the Governments of India.

ii) To collect industrial statistics in the organised sector by conducting annual surveys.

iii) To provide technical guidance for the conduct of crop estimation surveys and crop cutting experiments.

## b) The Central Statistical Organisation (C.S.O)

It was set up in the Cabinet Secretariat in May 1951, with the object of providing an agency for the co-ordination of a large variety of statistical data emerging from various sources in India. The C.S.O. performs many other functions in addition to co-ordination. In 1961, the status of C.S.O. was raised to that a department with the creation of a Department of Statistics inside the Cabinet Secretariat comprising the Directorate of NSS also as its subordinate office. Recently a computer centre has also been established under the Department statistics.

The divisions of the C.S.O : The C.S.O. performs many other functions in addition to co-ordination. It performs these other functions through the following divisions.

1. Industrial Statistical Wing.
2. Industry and Trade. Division including Village and Small industries unit.
3. Manpower Research Division
4. Methodology Division.
5. National Sample Survey Division
6. National Sample Survey Division
7. Planning and State Statistics Division

8. Population Division
9. Analytical Division
10. Family Living Survey Division
11. Prices and Cost of Living Statistics Division
12. Statistical Intelligence Division
13. Training Division.

The Industrial Statistical Wing is located at Calcutta and the remaining Divisions are at its headquarters in Delhi.

## Functions of the CSO

The main functions of the CSO are :

1. Co-ordinating statistical activities at the centre and the states.
2. Tendering advice to Central ministers and Government organisations on statistical matters.
3. Provision of national data to the United Nations and other International statistical organisations.
4. Regular publications
5. A graphical presentation of economic land social statistics.
6. Undertaking research in development of standards and methodological studies
7. Prepare National Income accounts.
8. Compiling and publishing Reports on Annual Survey of Industries.
9. Construction of Consumer Price Indices.
10. Organising and conducting training courses in official statistics.
11. Preparing estimates of population every year. The census of population in India is taken once in ten years. The estimates of population are needed year - wise to frame suitable policies in the different spheres. These estimates are prepared by C.S.O.
12. Organising statistical conference and committees and render secretarial assistance.
13. Taking up any work pertaining to collection of statistics on the request of any State Government or Central Government.

**Publications of C.S.O.:** The C.S.O. brings out a large number of publications. Some of these are :

1. **Statistical Abstract of India (Annual)**

It contains detailed statistical information covering practically all important aspects of Indian economy. Important among such items are; area and population, migration, national finance, income-tax, exchange, coinage and currency, banking and insurance, index numbers of price and wages, agriculture, industries, joint-stock companies, co-operative societies, mineral production, labour and trade, transport and communication. All India totals are provided in most of the tables for the last five or ten years.

2. **Monthly Abstract of Statistics**

This population provides detailed statistical information pertaining to various facts of the Indian economy for the month under review. The major heads of data include- population and vital statistics, employment, trade, industry and transport, banking, insurance, finance, prices, wages, index numbers etc.

3. Monthly statistics of the production of selected industries of India.
4. Statistical Pocket Book of the Indian Union
5. Annual Survey of Industries (10 volumes)
6. Sample Survey of current Interest in India (Annual)
7. Annual Survey of Industries - General Review
8. Monthly Bulletin showing production of selected Industries of India, and the index of Industrial production.
9. Statistical Newsletter (Quarterly)
10. DOCSTAT (Quarterly)
11. National Accounts Statistics - while paper (Annual)
12. Retail Price Bulletin-Consumer Price Index numbers for urban non -manual employees (Monthly)
13. Estimates of National Product (Revised Series) Annual.
14. Basic statistics relating to the Indian Economy (Annual)

Besides these regular publications, certain adhoc publications have also been brought out by C.S.O.

## Statistical Organisation in the States

Besides the statistical organisations of Centre, there are other organisations, bureaus and units at the state level. But statistical organisation in the state is of more recent origin than their counterpart at the centre. As former British Indian provinces, they were collecting certain statistics for the Centre on what were called Standardised Imperial Tables, on subjects like agricultrue, education, vital statistics, excise duty and so on. Before independence, there were in India, in addition to the provices, a very large number of princely states scattered all over the country and in different stages of social and economic development these princely states, with a few exceptions, had hardly any organisation which produced data of any kind beyond what was barely necessary for collection of revenue. After independence, these states were merged with the former provices, now known as the states of the Indian Union.

Since the War years, and particularly in the wake of the recommendations of the Inter-Departmental Committee of 1945, State Statistical Bureaus have been set up in most of the states either as independent statistical oranisations or as part of the combined economic and statistical set-up. In most of the states, the Department of economics and Statistics is responsible for the statistical data and information required by the states, statistical units are also associated to some of the state Ministries and other specific agencies.

The Conference of Central and State Statisticians held annually since 1961 (known as the Central Technical Advisory Council on Statistics since 1961) gave a further fillup to the setting up and strengthening of such organisations. Under a programme of central financial assistance under the Five - Year plans, the functions and the activities of all the statistical bureaus have also grown during recent years.

The State Statistical Bureaus are generally entrusted with the tasks of :

i) Co-ordination of statistics collected by different departments.

ii) Publication of statistical abstract assembling all essential serial statistics;

iii) Organising special enquiries and surveys;

iv) Liaison between statistical organisation of the centre and other states ; and

v) Statistical work relating to planning

There are, however, considerable differences in the different States Statistical Bureaus in the areas of their responsibility for collection of statistics. Agriculture, labour and vital statistics generally falls outside the scope of work of the State Statistical Bureau. Some of these Bureaus as these in the states of Uttar Pradesh, West Bengal and Maharashtra have been conducting a

number of socio-economic enquiries for collection of data required for formulation of policy in these states from time to time. In recent years many of the State Statistical Bureaus have joined the collaboration programme of the N.S.S. for conducting multi purpose surveys on a continuing basis.

An important development which took place in recent times in the state statistical system is the programme of central assistance extended to states under the Second and Third Five year plans for expansion of the statistical organisation and activities in the states. Some of the state statistical schemes for which Central assistance was extended are :

i) Strengthening of State Statistical Bureaus for planning needs.

ii) Setting up District Statistical offices;

iii) Setting up of Administrative Intelligence units for compilation of community Development / National Extension Service Statistics. and

iv) Training of Statistics Personnel:

Another noteworthy development in the statistics system in the states has been the appointment of a statistical assistant designated differently in different states as "Progress Assistant" "Statistical Officer" (Extension) etc; in each of the development blocks. Through this agency it had been possible to coordinate statistics of different types emanating at the block level. For supervising statistical work in the development blocks and also for co-ordinating the activities at the district level, District statistical offices were also set up on a phased manner.

The functions of the State Statistical Bureaus may be summed up as

i) Collection and Co-ordination of data.

ii) Liaison with C.S.O

iii) State Income estimation etc.

## AGRICULTURAL STATISTICS

Agricultural statistics in the basic level comprises of data relating to

1) Land utilisation, and

2) Area and production of crops.

## 1. Land Utilisation Data

### Content

Statistics for these consist of 1) total area, 2) classification of area, 3) area and crops irrigated and 4) area under different crops.

### Agency

These data are collected as part of land records by village revenue men (patwaris) and supervised by revenue officers.

The method followed is complete field enumeration on some of the states ie. in the temporarily settled areas of Punjab, U.P., Chennai etc., while in permanently settled areas of West Bengal, Bihar, Kerala and Orissa, this method is absent and what is followed is mere guess work by village headman or chowkidar; through in the recent years sample surveys are being used different crops during their regular rounds of surveys.

### Publications

i) These data are available on all India basis, from the Annual Indian Agricultural Statistics (Directorate of Economics and Statistics of the Ministry of Food & Agriculture). It is in two. Volumes : V.1 gives state-wise and V.2 district wise data. But these are available with a time lag of three years in respect of volume 1 and six years for volume 2. However, there is a mimeographed issue state figures of land utilisation with a time lag of one year only.

ii) The same data and with less lag are available for individual states separately in Season and Crop Reports which are issued annually by respective states.

iii) Summary statistics can be found in the following publications of the Directorate of Economics and Statistics.

a) Abstract of Agricultural Statistics (Annual)

b) Indian Agriculture in Brief (Irregular)

c) Agricultural Situation in India (Monthly)

iv) **NSSO:** Reports on some aspects of land utlisation surveys and crop cutting experiment used to give till 1970-71 their own estimates of data of land use under seven major food crops for the whole county, collected on sample survey basis; but after that they have started giving adjustment factors for the official estimates and have extended the coverage to all food and non food crops and all land uses.

These statistics now geographically cover much more extensive areas than before with the exception of non-reporting areas of Jammu and Kashmir, and parts of erstwhile NEFA.

The data are reliable in respect of temporarily settled areas where the system of field to field enumeration exists.

## 2. Area and Production

The statistics of area and production are available in respect of forest crops, plantation and some minor crops in the following publications.

1. Current estimates about area and yield are given for crops and plantation crops in the Monthly Agricultural Situation in India (DES Ag)

2. Annual data of area and production are given for all major and for some minor crops in the annual Estimates or area and production of principal crops in India (DES Ag, 2V) with V.1. for state figures; and V.2. for district figures, giving final estimates for the year, partially revised estimates for the last year and fully revised estimates for some years previous to that.

However, there is some difference in the estimates as given in Agricultural Situation in India and those given in Indian agricultural statistics. This is due to different approaches adopted in the two sets of statistics.

3. Season and Crop Reports issued annually by respective state governments.

4. Quinquennial report on the average yield per acre of principle crops (DES Ag)

5. NSSO : Reports of the consolidated results of the crop estimation surveys of principal crops 1955-56.

Yield statistics are collected through officials of the Agriculture and revenue departments of respective states with the guidance and supervision of NSSO staff.

The methodology has now been changed from the traditional method to stratified sampling survey of crop cutting experiments, and the same is being applied to more areas and crops. In the earlier method the estimates used to be subjective and the selection of plots was also not representative.

There are several other agencies giving yield estimates for tea, coffee, rubber, cotton and tobacco. The agencies are the Tea Board, Coffee Board, Rubber Board, Indian Central Cotton Committee, and the Central Board of Revenue. Their estimates, however, differ, from the official estimates because of difference in their scope, concepts and coverage.

NSSO has been giving annually their own independent estimates since 1955-56 covering the principal crops on the basis of yield rates as estimated from the sub sample series of the official crop cutting experiments.

## Cost of cultivation

Statistics of cost of production are collected for important crops only (wheat, paddy, jowar, bajra, gram, sugarcane, cotton, jute, etc.,) Data are available in the studies in the economics of farm management. These studies are conducted since 1954 by the Directorate of Economics and Statistics, Ministry of Food and Agriculture, in selected crop areas in various states, with the proposition to repeat such surveys in the same regions after an interval of ten years. Likewise ICAR has also made similar studies for sugarcane, cotton, jute etc., and their reports are available both in mimeographed and published forms.

NSSO has also conducted a sample survey to ascertain the cost of cultivation from the 5th to 7th round's. The results are published in its report on the same aspects of Cost of Cultivation Report No. 32 (for rural areas) and Report No.82 (for urban areas)

In fact, all these studies were simply adhoc surveys, with a limited geographical coverage, and were not repeated often enough.

However, recently a scheme has been initiated by the Ministry of Agriculture to collect data on cost of cultivation of principal crops in all the states on a continuous basis. For this purpose, a special cell has been created in the DESAg. collection of data will be made chiefly by the agricultural universities.

## Index numbers of production

The utility of index numbers of agricultural production lies in the fact that they show trend of production of food and other crops. They provide valuable indicators for planners in this country and are a guide for future projections.

There are four types of indices available.

i) Official series

ii) RBI series

iii) Non-official series

iv) FAO series

## Official series

The Directorate of Economics and Statistics of the Ministry of Food and Agriculture compiles these series of index numbers. These are published annually in its monthly Agricultural Situation in India (in January issue every year since 1949) The base year is changed to 1981-82. Coverage is extended to cereals, pulses, oilseeds, fibres, plantations and miscellaneous.

## RBI Series

These series are complied by the Reserve Bank of India and are available annually in December issue of RBI Bulletin since 1939-40.

## Non-Official Series

These series of index number are constructed by the 'Eastern Economist' and are available since 1939-40, its weekly journal in the last week of July each year.

Economic and Political Weekly and Indian Journal of Industrial Relations are some of the Journals engaged in constructing index numbers.

## FAO Series

FAO series include indices for India also. Base year is changed from time to time. Weighting system takes account of the world price relationship of agricultural commodities. These series are useful for making inter-country comparison. These are available in this Production Yearbook.

## Forest Statistics

The available data under forestry relate to :

1) area under forests 2) output and value of timber, fuel wood and other minor products 3) employment and 4) trade.

Data collecting agencies include Chief Conservators of Forests and Various Statistical Bureaus in states.

## Publications

1. This is the most comprehensive source in Indian Forest Statistics issued annually by DESAg. since 1947-48. Coverage is fairly comprehensive but there is a time lag of about four years. Again, output data are available for timber and fuel wood and not for minor products. State-wise output data of major forest products refer to areas, under state forest department. For other areas, data are not available. Employment data are available for public sector only.

2. Indian Agricultural Statistics (DESAg: annual) gives area figures only. But they do not tally with those of Indian Forest Statistics due to difference in concepts, and purposes for which data are collected in the two series, Geographical coverage and reference period also differ.

3. Annual administration reports of state forest departments.

## ANIMAL HUSBANDRY

### Livestock

Data about livestock, poultry and agricultural, implements are collected on complete census basis quinquennial, beginning form 1919-20 when the first livestock census, was conducted.

Collection of data rests with the states; in rural areas by revenue agencies and in their absence, by village chowkidar, and staff of the village panchayats; in urban areas by sanitary staff of the municipalities.

### Publications

Each State issues livestock census reports giving taluk-wise data. The information is consolidated and published by the DESAg. in the quinquennial Indian Livestock Census. This is in 2V, V-1 for state wise figures : V-2 for district data. Much delay is involved in the population. The provisional results are, of course, given in the Agricultural Situation in India. The fifteenth census was conducted in 1966, results of which where published in 1975. The last census was taken in 1972.

NSSO has been conducting, sample surveys of livestock since 1956, during its regular rounds, the results of which are given in Tables with notes on livestock products. These reports are useful for additional data given in respect of age, composition, breed, average lactation period, etc. of livestock. These reports also indicate the difference between the official census data and sample data.

### Livestock products

There are no regular data available for livestock products. However, some statistics are available. Marketing reports of the Directorate of Marketing and Inspection (DMI) which give estimates on suppiy and demand of certain products (till) now covering milk, ghee, other milk products, wool, meat, bones, bristles, and poultry eggs.

These estimates are, however not very reliable as they are not based on scientific method, Again, they refer to different periods.

ICAR is doing some work in this respect on the basis of sample techniques which it has developed to estimate to production of milk, eggs, meat and wool. Also, it has plans to conduct an all-India survey.

## FISHERY STATISTICS

### Tentative estimates

The chief source till recently, has been the report on the marketing of fish in the Indian Union issued by the Directorate of Marketing and Inspection. The last report was published in 1961, which contained tentative estimates of landing of sea fish, marketable surplus of fresh water fish, consumption, prices etc. The data are very much out of date now, and moreover, and not based on scientific method.

### Sample Survey

Sample surveys are being conducted regularly since 1950 by the Central Marine Fisheries Research Institute (Mandpam) through, the network of observation centres which collect data on total catches of individual species etc. The information is published in the quarterly Indian Fisheries Bulletin, (formerly Indian Journal of fisheries) of the Ministry of Food and Agriculture.

### Inland Fishery

Statistics of inland fishing are not satisfactory either and are not based on any scientific method. Data are collected by individual states, the returns of which are sent to the Fishery Development Authority for publication in the Annual Statistical Abstract of the Indian Union (CSO)

## DEFECTS IN AGRICULTURAL STATISTICS

The available statistics of agricultural sector have several defects which the users should be aware of :

### Basic Defects

The basic defects are due to inaccurate collection of data, inefficiency and dishonesty of the primary reporting agency and of the staff of the revenue and agriculture departments of various states.

### Gaps in Coverage

1. Geographical coverage of the data is estimated to be 89 % only.

2. Data in respect of certain minor items are lacking. Regular data on animal husbandry are not forthcoming. Reliable statistics of wages, hours of work, non-cash benefits etc. of agricultural labour are not available (see also labour).

3. No regular data are collected to estimate the cost of cultivation, except some occasional studies like NSSO: Report on some aspects of costs of cultivation which incorporates the result of survey made in the fifth to seventh Rounds October 1952 to March 1954.

4. Adequate data are lacking in respect of land holdings of agricultural labourers, except for certain adhoc surveys of the NSSO; the reports of the Agricultural Labour Enquiry and those of the Rural Credit Survey.

## Defects in Methodology etc

1. Absence of uniformity in concept and definition makes the statistics uncomparable over various regions, For example, there are different definitions of land-use classes in different states. Again, in certain areas, "harvest price" means average wholesale price at important centres during the harvest season, while in others, it is something else.

2. There is divergence in methods of data collection. For example, in temporarily settled areas, the primary reporting agency collects statistics; whereas in permanently settled areas, it is the village headman or even the village chowkidar to whom the work is entrusted. They do not take adequate interest in data collection which may mar the quality of statistics of such areas.

3. There are also gaps in processing and tabulation of data. Not all the information collected are tabulated due to absence of adequate processing system.

## Defects in Organisation

1. As already stated, there are serious drawbacks of primary reporting system. In the permanently settled areas of West Bengal, Orissa, Bihar etc; the statistics are collected by the village headman and the resultant data are not reliable. In the temporarily settled areas of U.P., Punjab, Tamil Nadu etc., the primary reporting agency is responsible for data collection, and the information is more reliable than in the former case. But still there are defects in this organisation also. The primary reporting agents are overburdened with work. They are untrained and are not adequately paid.

2. Further, lack of adequate supervision and absence of proper co-ordination among various agencies are responsible for defective organisational set-up of data collection.

## Time Lag in Publication

There is considerable delay in the publication of data (some times two or four years) as in the case of annual series of statistics. This makes the data lose their practical utility.

# TRADE STATISTICS

## INLAND TRADE

### Rail and River Borne Trade

The Director General of Commercial Intelligence and Statistics (DEGIS) collects the data of rail and river borne trade from the monthly returns submitted by the railways and steamer companies. The returns are based on the invoices.

The statistics relate to movements of selected commodities between thirty four trade blocks into which the country is now divided.

The information is available in the annual publication of accounts relating to the Inland (rail and riven borne) Trade of India (DGCIS 1933) : previously in the Annual Inland trade of India (DGCIS. 1923 : discontinued there after). There is a gap between the year 1924 and 1934 for which inland trade data are not available. The time in publication is about one year. The gaps in coverage are that;

i) not all commodities are covered by these statistics but only selected ones.

ii) parcel traffic is not included.

iii) trade by river is incomplete since the present statistics cover invoices of simply one steamer company. However, the Ministry of Transport and Shipping is contemplating to collect data giving full coverage.

iv) trade by road is completely left out which, by no means shares a less significant portion of the total volume of the country's inland trade.

v) value figures are not shown., only the quantities are given because this much only is obtained from the rail invoices and

vi) net quantities are not directly available, the quantities shown are gross of packing.

### Coastal Trade

Statistics of coastal trade are based on daily returns which are prepared from bills of entry by customs house and are submitted to the DGCIS. Data relate to movements (external and internal) of merchandise and treasures within and between twelve maritime blocks of the country and recorded in terms of both value and net quantity of trade handled at different sea ports of the country.

The information is available in monthly accounts relating to the coastal trade and navigation of India (DGCIS, 1866/67-1992): contained as Statistics of the Coastal Trade of India (DGCIS, April 1932/-) The latter was previously issued monthly, from April 1932 in March 1962; but now published annually. The time lag in publication is about four years. The gaps in coverage of data are follows:

1. These statistics exclude postal articles, passenger baggages, and transhipments of merchandise or treasure at a customs port.

2. The data now include merchandise of Indian origin only since April 1964; previously foreign merchandise also were included.

## Road Borne Trade

No regular data are available for inland trade by road. However, there are schemes to collect the required data, and recently the Planning Commission has approved such schemes whereby, the statistics will be collected by various state governments on sample survey basis at least once during a plan period. The schedule for operation of the survey has been finalised in the Guidelines for conducting goods traffic survey by road (CSO, mimeographed paper).

## Air-Borne Trade

No data are available for airborne trade, except to the extent of total freight carried which is provided in the Anual Basic Statistics of Air Transport in India (Directorate of Air Transport, 1961) and previously in the Hand book of Civil Aviation Statistics, 1959-1960 issued by the same Directorate. The statistics provided in these publications do not show commodity-wise distribution of goods traffic.

## Distributive Trade

In the field, no regular data are available. However, there have been a few adhoc surveys in the household sector which were conducted by the NSSO in its various rounds, from seventh to tenth rounds, fifteenth to twentieth rounds and in the twenty fourth round, provide some light in this case.

In its first round, NSSO covered rural areas only (its report no.1) The scope was somewhat general in nature. The seventh to ninth rounds of survey (October 1953 to November 1955) attempted to give some broad information on the household trade (its report no.24), which covered whole of the country except Jammu and Kashmir, and Andaman and Nicobar Islands. The NSSO Report no.14 gives data for the tenth round (December 1955 to may 1956); its Report no.99 gives data for the fifteenth round (July 1959 to June 1960); and Report no.218 for data after tenth, the

geographical coverage has been the whole of India including Jammu and Kashmir, except Andaman and Nicobar Islands; and in the twentieth and twenty fourth rounds, the scope and sample size were further extended.

## Foreign Trade

### DGCIS series of data

The DGCIS compiles and publishes statistics of foreign trade of India on the basis of Daily Trade Return which are prepared by the customs authorities at seaports, airports and at the land customs station on the basis of shipping bills and bills of entry respectively for exports and imports.

The statistics of foreign trade include quantity and value figures for the month and for the comulative period in respect of exports, re-exports, and imports of merchandise and treasures from and into the country. These are recorded in detailed tables by commodity under various groups, divisions and sections of the Revised Indian Trade classification and by country. Also, summary data of trade are available and country with corresponding figures in the previous year.

However, trade with Bhutan and Sikkim is not included. Trade with Nepal is included - land trade statistics from April 1964, and those of airborne trade since January 1957.

### Publications

These statistics are regularly published in the Monthly Statistics of the Foreign Trade of India (MSFT 1) by the DGCIS, since January, 1957, in 2V, with V. 1 for exports and re-exports; and .2 for imports.

Previously, these data were available in the DGCIS publications, Accounts relating to the foreign trade (Sea and airborne) and Accounts relating to trade of India by land with foreign countries upto 1952, Accounts relating to the foreign trade (Sea, air and land) and navigation of India from 1953 to 1956 and the accounts relating to the foreign trade and navigation of India upto April 1956. The time lag in publication of the present serial is about six months.

### Coverage

i) Transhipment trade is not included in these statistics, but in the supplement.

ii) The statistics do not include non-dutiable personal baggage trade in ships, aircrafts and combat defence equipment.

iii) These data also do not include trade through postal parcels; and transactions in treasure.

iv) Consignments from India to diplomatic missions and armed forces and consignments from abroad to their counterparts in India are, however, excluded.

## Other Publications

1. A good deal of supplementary data are given in the supplement to monthly statistics of the foreign trade of India (DGCIS January, 1957 - quarterly).

2. The data contained in the MSFTI are verified and revised at the year and are given in the Annual statistics of the Foreign Trade of India; by countries (DGCIS, 1959-60);

3. Annual statistics of the foreign trade of Indians by customs Zones (DGCIS, 1957-1959), provided Zonal distribution of data of the country's foreign trade for the said period. This is now being given in the main publication, viz. MSFTI.

4. Statistics of foreign and coastal cargo movements of India (Ministry of Commerce, 1965-annual). This was earlier published as Statistics of the maritime navigation of India (DGCIS, January 1957-1965, annual). They contained navigational data of India's foreign trade.

5. For over-land trade with Nepal, Sikkim and Bhutan, information is given in the weekly. Indian Trade Journal which also provides statistics on bilateral and other trade agreements.

## RBI Series of Data

The Reserve Bank of India also provides, in value terms, statistics of exports and imports on the basis of balance of payments data.

These data are published on quarterly and annual basis under the Balance of Payment series in the RBI bulletin. The classification of data is broad with 24 commodity groupings.

The RBI's balance of payment data differ from the foreign trade statistics compiled by the DGCIS.

## The major Differences are

i) Trade between India and Nepal does not get included in the balance of payments data since there are no exchange restrictions while on the otherhand, this is now recorded in DGCIS data, though not fully.

ii) Coverage of commodities also differs. The RBI data cover only these commodities which enter in the foreign exchange transactions. While DGCIS data cover all commodities which enter or leave the country, the RBI data include purchase of all defence equipments, while the DGCIS data record only non-cambat items of defence.

iii) Recording of imports under deferred payments are not fully done in RBI series; whereas DGCIS records them fully.

iv) There is difference also in the system of valuation in the two series of data. The value of exports as recorded in the customs data is always f.o.b., while in the balance of payments data, it may be f.o.b., c.i.f., c.f. etc.

v) The RBI data take account of the value added by repairs and improvements, while the customs data record the full value of the equipments repaired.

vi) The customs data record exports as and when these are consigned and imports when they arrive; but the RBI data record the imports when payments have been made and in the case of exports when the goods have been shipped.

vii) The RBI compilation is concerned with earnings payments of foreign exchange in different currencies and from various regions of the world, without giving minute classification on commodities exported and imported nor quantities thereof.

## Gaps in Foreign Trade Statistics

1. Trade between India and Nepal are just approximates, only trade through some eighteen routes are recorded out of 300 overland routes. For other routes, coverage is given by escalating the data available for eighteen check posts. This much is available from DGCIS compilation, but nothing of this in RBI data.

2. In respect of imports, all items are not recorded in the first instance, as in the case of foodgrains and petroleum, where under the system of 'Notes Pass', some delay occurs in the receipt of all details at the DGCIS and accordingly the corresponding data are deficient. However, revised sets of data are provided annually in the Annual Statistics of the Foreign Trade of India.

3. DGCIS data are deficient in the case of exports on consignment basis where adjustment for short and shut-out shipments has to be done. The customs data used to give cover for such cases previously when it was the practice to record exports on the basis of duplicates of the shipping bill which forms the basis of recording of exports at the DGCIS and no corrections are made later on in such cases of short shipments etc.

4. The value of export of films is at notional price and does not represent the actual transaction value.

5. Unauthorised imports are, by their nature, absent from the available data.

6. Non-cash imports are not covered in the RBI series of Data.

7. Commodity-wise details of aid-financed imports do not get fully reflected in RBI's balance of payments data.

8. Because of the absence of a common scheme of classification, there are discrepancies in the commodity-wise balance of payments data and the DGCIS statistics.

9. No statistics are available either on the mode of finances of import items or about their end uses (like product imports, maintenance imports, etc.)

10. There is no provision in the data collected at present to know the share of individual sets in the country's foreign trade. However, this point is being pursued by the Government.

11. One of the unexplored sources of foreign trade statistics is that "the bills of entry contain valuable information on c.i.f. prices, f.o.b prices, freight component etc. on various commodities. But the authorities never care to compile these data.... Researchers in the field of international trade pretty well understand the great value of the price data in analysis of trade problems". There is need for tabulating data from DTR reflecting the "end use structure of imports by commodities.... which will be useful in evaluating the import licensing decision etc."

12. The Monthly Statistics of the Foreign Trade of India (MSFTI) does not give aggregates for the generic group of commodities like handicrafts or engineering so as to watch the trend of trade in it. What it gives is at the micro level classification at the point of seven and three digit groupings, showing against each country, monthly and cumulative figures of trade in that commodity. Even at seven digit commodity level, the absence of comparative figures for previous months and years makes it difficult to know the progress or otherwise of the trade in that commodity when consulting simply one single issues of the MSFTI. Yet another gap which is pointed out is that the country-commodity figures are not available in MSFTI are also equally required by the policy makers besides the commodity country figures.

## INDEX NUMBERS OF FOREIGN TRADE

### 1. DGCIS index

This is available in the quarterly supplement to the monthly statistics of the foreign trade of India.

The index numbers refer to both exports and imports and under each volume and unit value.

### RBI Index

This is available in the RBI Bulletin in its January issue every year. The indices are constructed on monthly, quarterly and annual basis.

The RBI index which represents selected items from all routes sea, air and land is more representative than those of the DGCIS, which does not include items of trade by land.

# INDUSTRIAL STATISTICS

Industries in India can be divided into two categories: large scale and small scale.

## Organised Sector : Large Scale

The organised sector comprises of all manufacturing units which employ ten or more workers when using power, and also those which employ twenty or more workers when not using power. Of these units which employ fifty or more workers when using power or those which employ 100 or more workers while not using power are called the large scale factory establishment. The rest constitute the small scale establishment in the organised sector.

The statistics for the organised large scale sector are available from.

1. Census of India Manufactures (CMI)
2. Sample Survey of Manufacturing Industries (SSMI)
3. Annual Survey of Industries (ASI)
4. Monthly statistics of production of selected industries in India (MSP)

## 1. Census of India Manufactures(CMI)

Collection data in CMI was based on statutory returns from individual units which were submitted to the respective state statistics authorities. Thereafter, the information was consolidated, processed and published by the Directorate of Industrial Statistics. The Directorate was set up for this purpose in 1945, and was later merged with the CSO in 1957.

Statistics relate to value and volume of production, consumption of raw materials and power, labour employed, capital invested and value added by manufacturer.

## 2. Sample Survey of Manufacturing Industries (SSMI)

Collection of data under this programme was the result of recommendations of the National Income Committee.

Data were available under SSMI only to the extent of units employing ten or more workers with having power, as well as excluding, however, the defence establishments and railway workshops.

The two programmes of CMI and SSMI continued more or less side by side, neither of which could possibly replace the other one. They served two different purposes the SSMI provided blown up estimates for the whole population; the CMI for the actual reporting factories only. The

former was intended to provide the basis for estimates of national income of the manufacturing sector, while the CMI was meant for the collection of industrial statistics for the specified factories in existence.

Data given in these two sources, though much out of date now, should be used with this view in mind, irrespective of the fact that the coverage and quality of the SSMI data were better than those of the CMI.

## 3. Annual Survey of Industries (ASI)

Both CMI and SSMI have now been replaced by the Annual Survey of Industries (ASI) since 1959. The first ASI was conducted in 1960 for the reference year 1959.

The entire field work is in the charge of the Directorate of National Sample Survey. The processing and tabulation work is divided between the Industrial Statistics Wing (ISW) of the CSO for census sector and the Indian Statistical Institute (ISI), Calcutta, for the sample sector.

The ASI covers the entire factory sector in the country coming under the Factories Act, 1948. Since the Act does not apply to Jammu and Kashmir, the factories in that state are covered on voluntary basis.

The ASI includes factories employing fifty or more workers when using power and those employing 100 or more workers when not using power, on complete census basis, taking some 13,000 units; and factories employing ten to forty-nine workers when using power and those employing twenty to ninety-nine workers when not using power, on one-third probability sample basis taking some 36,000 units The census sectors alone covers four-fifths of the total output of the two sectors.

### Items of Information

Data are collected on all -India and state level for the following items:

1. Capital structure
2. Employment and wage data, arranged by skilled and unskilled labour, industrial relations training etc.,
3. Inputs: consumption of raw materials, fuel etc.,
4. Output : Quantity and value of manufactures;
5. Stocks
6. Installed capacity

7. Power equipment

8. Sales ; and

9. Value added by manufacture.

Information collected relates to the operation of factories in the previous year. Reference year is calendar year. In most cases response is satisfactory and had increased to 98% in 1965 from 94% in 1959.

## Publications

Results of the ASI in census sector are published by CSO and for the sample sector by the National sample Survey Organisation. But there is a long delay in the processing and publication of the data. Till date, detailed reports are lacking for years 1967 onwards. ASI data are now being tabulated at state level; district -wise tabulation is also being attempted by some states. This is confined to census sector only.

## Gaps in Data

i) More details of capital structure are desirable for estimation of capital formation in the industrial sector, which are lacking in the available data.

ii) Installed capacity data are collected under ASI since 1959. But these are not tabulated or published. This has to be filled up from the Monthly statistics of production of selected industries of India (MSP), But the coverage of ASI is more comprehensive than MSP.

iii) Sample part does not give details of inputs and outputs like the census part.

iv) Classification of industries is more detailed in census part than in sample part.

v) Inordinate time lag in publication of data is not at all helpful.

## 4. Monthly Statistics of Production of Selected Industries in India (MSP)

Till now we have discussed collection of statistics under statutory rules. These are regularly available since 1946 successively under CMI, SSMI and ASI, On voluntary basis also some statistics are being collected, as it were, to make up the time lag involved in the availability of annual data under CMI, SSMI or ASI.

It is the CSO which consolidates and publishes the information collected on the basis of returns from Director General of Technical Development (DGTD) and fourteen other official agencies. The DGTD itself is responsible for information collected in respect of over eighty percent of the products.

These statistics are regularly available since 1947 in the publication of selected industries in India (MSP) which was previously by the Directorate of Commercial Intelligence and Statistics, Calcutta and now by the CSO. This publication gives, by state and by industry put data of about ninety major industries of the country along with statistics of installed capacity stock figures and index of industrial production. The coverage of industries is only partial unlike ASI, because of the voluntary nature of the system of data uncomparable over time and by industry. However the geographical coverage in MSP is the whole of India minus Jammu and Kashmir and the data are useful to some extent.

## Gaps in Data

i) Capacity data given in MSP are based not on the actual working conditions in the factory or technical feasibility considerations but on the basis of number of shifts per day and number of working days in a year. It has been observed through several studies that the MSP data of capacity far exceeds the figures of production in a large number of cases, while in others, the capacities were underestimated.

ii) In most cases, MSP data relate only to larger establishment having more than Rs.7.5 lakh as capital in plant and machinery.

iii) The coverage of MSP data for a given product in relation to its total production in the economy, is not clearly known and this coverage seems to be changing over time.

iv) The product coverage also is not complete. The DGTD product coverage is highly incomplete. However, in such cases where those fourteen agencies collect data and have statutory powers to call on returns, the coverage is more or less complete. This is the case with cotton, jute iron and steel and sugar industries.

v) Non-response is considerable in most of the cases as the filling of returns is voluntary.

vi) No information is given as to reasons of under-utilisation of installed capacity.

vii) Time lag, too is considerable.

## Other sources

Alternative sources to MSP may also be consulted mainly because they have lesser time lag. These are the following :

1. CSO has been issuing, in a mimeographed form, the Monthly Bulletin on index of industries and their indices.

2. The Annual report of the Director General of Technical Development (DGTD) contains production data of selected industries as covered by it and submitted to the CSO for publication in MSP.

3. Similar publications of other official agencies which collect data for the MSP can also serve the purpose, and these have a lesser time lag. These are

a) India, Director General of Mines Safety; Monthly coal bulletin.

b) India, Ministry of Food and Agriculture: Regional Office of Jute Development; Monthly Jute bulletin.

c) India, Office of the Textile Commissioner : Indian Cotton Textile Industry (Annual) and Indian Textile Bulletin (Monthly)

d) Indian Sugar Mills Association, Kanpur:

e) India, Tea Board: Tea statistics (Annual

f) India, Coffee Board: Coffee statistics (Annual)

g) India, Rubber Board, Rubber Statistics (Annual)

h) India, Department of Iron and Steel : Iron and Steel; Iron and steel control monthly Bulletin.

i) India, Central Water and Power Commission : Public electricity supply; all India statistics (Annual)

j) Petroleum Information Service: Petroleum Hand Book (Annual)

## Small Scale Sector

This sector consists of all manufacturing and processing units in the household and non-household level whose capital investment does not exceed 0.75 million, including repair and maintenance services.

## Current Data

In respect of small scale manufacturing units in the non-household factory sector, data are now being collected through ASI on a phased scheme, covering in the first instance, priority groups of industries, like metal products and machinery, chemicals, textiles and other industries. Data are being collected on Census basis since 1966.

The field work is conducted by NSSO; processing and publication of results by the ISW of CSO

Summer reports are brought out covering metal products industries. This is available in the publication Annual Survey of Industries; Small scale industries sector Data relate to capital, labour wages and emoluments, inputs of fuel and material, output of products and by-products, depreciation and value added by manufacture. Industry-wise and state-wise data are given. However, the 1966 report does not give as many details as the 1967 report.

## Past Data

The past data about this sector is available only in respect of total employment and number of units, and even there not directly but by the method of residue, i.e., by deducting the figures available for large scale sector (In large industrial establishments in India), from the total employment figures available in the decennial censuses.

## Cottage Industries

No regular and continuous statistical series are available for this sector. However, some adhoc statistics are available on national level from NSSO reports and those of the individual cottage industry boards and from the National Small Industries Corporation. On the state level, some states have conducted adhoc surveys at different times, but they are of little use.

## NSSO Reports

The NSSO Surveys are not regular. Also, these are confined mainly to household enterprises, except in a few rounds.

Such surveys were conducted from the seventh to tenth round (October 1953 to March 1954), and in the fourteenth round (July 1958 to June 1959). The results are available in its report on 19,21,42,43, and 94 with the title. Tables with notes on small manufacture; rural and urban household enterprises smaller than registered factories.

The data collected in these surveys relate to fixed capital, employment value of input and output and value added. These data are available for the sector as a whole and for certain industries individually.

## Cottage-Industry Boards

There are a number of cottage industry boards in the country like the Handloom Board, Sericulture Board, the Khadi and Village Industries Commission. They collect some data on production in their respective fields which are published in their annual reports etc.,

Statistics of production from handlooms are available in the Indian Textile Journal (Office of the Textile Commissioner); for silk, in the silk journal (sericulture Board); in respect of khadi cloth,

in the Annual Report of the Khadi and Village Industries Commission. For handicrafts, regular data are not available, expect that the Handicraft Board, collected data from individual states and published them in its statistical bulletin of March 1960. The information relates to number of units, employment and value of output in each industry.

## Census of India

In the Censuses of India 1961, and 1971, several monographs on village surveys and rural handicrafts have been issued in almost all the state series and several others are yet to come which will form a good repository of statistical data for this sector.

## National Small Industries Corporation

Some data are also available from surveys conducted by the NSI Corporation. The results are published in its Analysis and Planning Reports. The data relate to capital, employment, raw materials, output etc.

Under the Fourth Five Year Plan, and All-India Surveys of all units in the country, which employ five or more workers specially in urban areas, are being conducted on a uniform pattern in all the states. A great deal will be available when the results are published.

## Index of Industrial Production

The current index which comprises both monthly and annual series of indices of industrial production is available in monthly statistics of production of selected industries of India (CSO, 1948) and with much loss time lag-in its mimeographed publication Bulletin on Index of Industrial Production (CSO , 1963, Monthly)

The index has the base year 1950 and is composed of 324 items for monthly series and 446 items of annual series. Weights are assigned on the basis of value added by manufacture according to ASI 1960 both census and sample parts : and the value added in the case of mining from Indian Bureau of Mines, and in respect of electricity from figures of the Central Water and Power Commission.

## Past Index

i) The construction of official index numbers of industrial production was started by the office of the Economic Adviser, Ministry of Industrial Development and Company Affairs. The index had base year 1937. Covering fifteen industries (Iron and steel, cement, chemicals, paper, matches, paints, cotton textile, jute manufacture, wheat flour, sugar, petrol, kerosene, coal, electricity; distilleries and breweries) This series was discontinued in 1949.

ii) Revised index was then computed by the Directorate of Industrial Statistics, which was started from January 1949 with the base year 1946. It was published in Monthly statistics of production of selected industries of India from January 1950 till April 1956. It took into account thirty six commodities in twenty industry groups.

The above series of index numbers have undergone several revisions. The present base year for general index of industrial production is 1993-94.

### Defects

The user should be aware of the limitations of the present series of index numbers. The weighting is based on figures obtained from a wide variety of sources, and as a result, it does not reveal uniform quality. Secondly, the production figures that form the basis of weighting are available only in the case of large scale units covered under the ASI. Other units which have a great deal of share in the country's total production are left out in this way.

## Other Sources of Industrial Statistics.

1. Journal of Trade and Industry, published by the Directorate of Commercial Intelligence and Statistics, Calcutta, provides some data on production, consumption. Export and import, besides feature articles dealing trend of progress of industries.

2. Reserve Bank of India Bulletin, (weekly) also contains figures of industrial production and the index number.

3. Statistics of the Working of Capital Issues Control, (Quarterly) issued by the Ministry of Finance, provides, on a quarterly basis, data of capital amount granted by the Government for issue of fresh capital by joint stock companies.

## Model Questions

1. Describe the availability of the various types of statistical facilities, organisation and data available in India.
2. Describe the availability of trade statistics in India.
3. Write a note on the existence of industrial statistics in India.

## LESSON - 15

# LABOUR STATISTICS

Labour statistics constitute one of the most vital economic indicators in a country. From the point of view economic planning these are of utmost importance. Formulation of labour policies, planning, production, review of employment situation, implementation of social welfare measures etc., are not possible without adequate and accurate labour statistics.

A good deal of data are collected and published in India on various aspects of labour economics. The main aspects covered include; employment, unemployment, hours of work, absenteeism, industrial injuries, trade unions, industrial disputes and industrial relations; and wages and earnings.

## EMPLOYMENT

## RETROSPECTIVE STATISTICS

### 1. Census Reports

Chief course of retrospective statistics is the devennial census reports of population.

Data are available for employment by industry, by occupation and by regional distribution. However, the information given may not be easily comparable over time, since there have been changes in definition, concepts and scope from one census to another.

### 2. NSSO Reports

Reports of NSSO relating to various rounds of surveys on employment, unemployment and underemployment are other sources, specially its Reports on Employment and Unemployment (Ninth round May-Nov. 1955)

## Current Statistics

Current data of employment are regularly collected in the organised sector for 1) factories 2) mines, 3) plantations, 4) transport and communications, and government services both state and central.

There are a number of agencies which collect data. The collected statistics are passed on to the Labour Bureau where these are consolidated and published annually in its Indian Labour Statistics. Also similar statistics are available in Labour Bureau's Indian Labour Journal (monthly) in Annual Survey of Industries, and in the Annual Statistical Abstract of the Indian Union both of the CSO.

## Factories

The Chief Inspectors of Factories in various states collect data of employment for factory sector which are submitted to the Labour Bureau; Simla. Response is not always good. In defaulting cases, the average daily employment is estimated on the basis of factory inspection

reports. The scope of the operation is limited under the Factories Act 1948 to units employing ten or more workers, when using power and also those employing twenty or more workers, when not using power.

The data provide average daily employment shown by industry, by state, and also separately for female labour. These data are published annually by the Labour Bureau in Indian Labour Statistics.

Data are also collected under the ASI since 1959 where figures for average daily employment are available separately for men, women and children for about 200 industries as against only twenty nine groups of industries under CMI previously.

### Mines

The collecting authority is the Director General of Mines Safety under the Mines act 1952. The statistics relate to average daily employment of wages earners, salaries, staff, foremen and apprentices in mines and quarries, shown for all major mines separately.

These data are available in the Statistics of Mines in India (Director General of Mines Safety. 1995 Annual in 2V); and for data previous to 1952 in the Annual Report of the Chief Inspector of Mines and also in Indian Labour Statistics.

These data are confined to important mines and quarries covered under the Mines Act of 1952 amended in 1959. Excluded are the category of mines exempted by the Central Government.

### Plantations

The Directorate of Economics and Statistics compile data of employment for coffee and rubber plantations. For the latter is another source, the Annual Tea Statistics (Tea Board, Calcutta, 12955) which gives district-wise employment figures. The employment data include all workers in and outside the plantations.

### Railways

Collection of data is done by the Railway Board. The Statistics relate to number of persons employed as on 31st March every year covering gazetted and non-gazetted staff on open lines and construction including the staff of the Railway Board. All railways are covered. The data are given in the Report by the Railway Board on Indian Railways, V2: Statistics.

### Posts & Telegraph

Statistics of number of persons employed as on 31st March each year are collected by the Director General of Posts and Telegraph and supplied to Labour Bureau. The data are given in Indian Labour Statistics. The information is confined to non-gazetted staff only.

### Motor Transport

Data are now being collected under the Motor Transport Workers Act 1961 from all companies employing five or more workers. These statistics have been published for the first time for reference year 1968 in Indian Labour Statistics 1971.

## Banks and Insurance

For banks, no official data are available on regular basis. For Insurance, data are available in Monthly Abstract of Statistics.(CSO)

## Commercial Establishments

Only partial statistics are available in Indian Labour Statistics, relating to a few states and there too, in respect of some selected urban areas where the shops and commercial Establishments Act is applicable.

## Coverage

i) All units in public sector : central, state, semi-government and local bodies, collect quarterly; data

ii) All units in the non-agricultural, private, organised sector employing twenty five or more workers, collect quarterly; data and

iii) All units in the same sector (2 above) employing ten to twenty four workers, collected quarterly data on voluntary basis in 316 districts and all units employing five to more workers on biennial basis.

It covers the whole of India except Goa, Daman & Diu and Laccadives. The statistics relate to total employment shown state-wise and zone-wise. Employment in public sector, classified according to different groups as central, state, quasi /Government and local bodies employment in selected industries, shown separately for public and private establishments employing five small private establishments employing five to nine workers with figures given state-wise and by industry.

These data are regularly published on all India level, in Quarterly Employment Review (DGET.1961), and Employment Review (DGFT 1961 Annual & quinquennial). For the private non-Agricultural sector where the units employ five to nine workers, these statistics are given in the Employment is smaller establishment, biennial survey; private sector (DGETM 1964)

On the state level detailed statistics by the district and by industry are available regularly in Employment Review issued by respective state Employment Directorates and published either biennially or annually.

## Occupational & Educational pattern of Employment

Studies on occupational distribution of employees have been taken up on regular basis under the EMI programme. They cover separately public sector and private non-agricultural sector where the units employ ten or more workers. Data are collected on all India level covering public sector in one year and all the private sector in another year alternately since 1966. Data on educational pattern also are being collected starting with the public sector in the first instance.

On the state level similar information is contained in publications of state employment directorates which are being brought out since 1960/61 either annually or biennially.

## UNEMPLOYMENT

For statistics on unemployment there are three sources, but none of them provides a continuous time series of reliable data. The sources are as follows:

### 1. Census of India

Data on unemployment was collected for the first time in the 1961 Census. State-wise distribution of unemployed persons is given. Reliability of data is doubtful, because of the gigantic size of the census operation.

### 2. Employment Exchanges

The employment exchanges in the country located mostly in urban areas, serve perhaps the only source of giving continuous information about unemployment.

The statistics which the employment exchanges provide relate to the number of persons registered at the end of each month (classified occupation -wise) and the number of persons who got employment.

These statistics are available monthly in Indian Labour Journal and annually in Indian Labour Statistics.

But this source has some serious defects, such as : incompleteness of information, multiple registration, absence of correct information about the number of persons employed out of the registrants; the voluntary nature of the rules for registration which leaves out a big chunk of unemployed persons. Moreover rural areas are not covered at all.

### NSSO surveys

Sample Survey conducted by NSSO provides data on unemployment. First such survey was conducted in 1953. The results were made available in its Report on preliminary survey of urban unemployment (NSSO report No.8, 1956). It was confined to towns with population of 50,000 and above, excluding the four metropolises. Another survey was conducted in 1956, for the city of Calcutta (NSSO report No.17, 1959)

Surveys covering whole of the country for employment and unemployment were started from the ninth round (May,1995 to November, 1955) on a continuous basis. The results of this round were contained in NSSO report No.16, 1959. This is continuing. However, the areas of NEFA, Ladakh, Andaman and Nicobar Islands and Laccadives are excluded from these surveys.

The coverage of rural areas is not always complete. On all the rounds; sometimes rural areas are covered in and at other times urban areas alone are covered.

### Wages and Earning

Wage data for the organised non-agricultural sector are collected under different acts by various authorities and are published by them in their respective annual reports etc. Consolidated

data are also available regularly in Indian Labour Statistics (Labour Bureau, 1959) showing average per capita annual earnings in various sectors by industry and according to various components of wages.

## Factories

Data are collected under the Payment of Wages Act by State Labour Department from individual factory establishments and are submitted to the submitted to the Labour Bureau. Statistics of wages for factories are also collected under ASI since 1959 (Previously under CMI& SSMI, 1946-1958)

The data collected in the former case cover factories with ten or more workers (with power) and those with twenty or more workers (without power). In the latter case, larger units only are included.

## Mines

Statistics of wages and earnings in the mining industries are collected by the Director General of Mines Safety. The wage data for workers in mines other than coal mines refer to average daily earnings for the whole year instead of average daily earnings during December each year, as was the position upto 1960. For coal mines, the average daily earnings relate to each month of the year.

These data are available in Indian Labour Statistics (Labour Bureau) and in the Statistics of Mines in India (Director General of Mines Safety).

## Plantations

For tea estates in Assam, the information is collected by the Controller of Emigrant Labour on the basis of returns from individual tea estates. This is given in the Annual Report on the working of the Tea District Emigrant Labour Act 1932, in addition to Indian Labour Statistics. The data relate to tea workers only and do not include clerks, servants, carpenters, etc.

For tea estates in West Bengal, the Labour Commissioner supplies the data which are published in Indian Labour Statistics Journal. The data relate to workers with annual earnings of less than Rs.400/-

For other plantations like coffee, rubber etc. various state authorities compile the data under the Payment of Wages Act and the Reports of the Working of Payments of Wages Act. However, consolidated data are also given in Indian Labour Statistics Journal.

## Railway & dock Labour

For Railways, the average annual earnings for class III and IV staff are available in the Annual Report of the Railway Board.

For dock labour, data of average monthly earnings are given in the Annual reports of the respective Dock Labour for Calcutta, Bcmbay, Madras, etc. Summary statistics are available in Indian Labour Statistics Journal.

## Unorganised Sector

The unorganised sector in this case is that the sector of the industry which is not covered by the Payment Wages Act have been applied. This excludes the agriculture sector. Chief Labour Commissioners in individual states collect wage data in this case.

These statistics are published in the Annual Report on the working of the Minimum Wages Act, 1948 (in states) issued by individual state government and are also consolidated in Indian Labour Statistics (Labour Bureau)

## Adhoc Surveys

Besides serial statistics, there have been a few adhoc surveys which are aimed at obtaining adequate data for the construction of index numbers of wage rates for various industries. These include collection of data on wage components, wage differentials and distribution of workers by wage level.

i) The most important survey that was conducted at the official level was the Occupational Wage Survey conducted by the Labour Bureau. The first sample survey was taken in 1958-59. Covering forty four major industries in the manufacturing, mining and plantations sectors. The results were published in Occupational Wage survey : General Report, 1958-59 and various industry reports in 5 Volumes (Labour Bureau, 1963-1967).

The data collected relate to "industry-wise distribution of workers by employment status, by methods of wage payments, by wage period and by sex : wage rates - minimum and maximum : average daily earnings, by components, for time rated and piece-rated workers, and for men, women and children, and distribution of workers by level of earnings".

The second Survey was done in 1963-65 with more or less the same scope and coverage. The summary report gives on all India basis, the data regarding employment, average wage rates, employment status of workers, method of payments, payroll earnings and incentive schemes.

ii) Another adhoc study was made by the National Commission on Labour. The results were published in the Statistics of the Selected Manufacturing Industries (Ministry of Labour, 1969, 3 parts.)

## Agricultural Labour

Wage data for agricultural labour are available annually on continuous basis since 1920 in Agricultural Wages in India (Directorate of Economics and statistics. Ministry of Food and Agriculture). But employment statistics are not available except for the reports of the Agricultural Labour Enquiries.

## Agricultural Labour

The Labour Bureau has conducted comprehensive agricultural labour enquiries to collect data on employment, wages and earnings, cost of living and independence of agricultural labour.

The first Agricultural Labour Enquiry was conducted in 1950-51 on sample basis in 813 selected villages covering about 11,000 households. It was in three stages:

a) The results of the first stage enquiry were published in the Agricultural Wages in India, Vo I & II (Labour Bureau, 1952). The data collected relate to (1) economic condition of the sample villages (2) system of land tenure, (3) methods of land use (4) wholesale and retail price of selected items of consumption and (5) prevailing rates of wages.

b) The results of second stage enquiry were published in Rural Manpower and Occupational Structure (Labour Bureau, 1954). The data collected refer to size, composition, occupation and earnings of all families in the sample villages.

c) The third phase of the enquiry was devoted to employment, unemployment, wages and earnings, consumption expenditure and rural indebtedness of 11,000 sample families. These data were published in Agriculture Labour Enquiry 1950-51, Report on Intensive Survey of Agricultural Labour. ( Labour Bureau, 1955) in 7 V, with 5.1 giving all India summary results, and the remaining 6 Volumes giving state-wise reports.

(2) The Second Agricultural Labour Enquiry was held in 1956-57 covering 3676 villages and 28560 sample households. The results were published in volumes in the Agricultural Labour in India. Several Reports on the agricultural labour enquiry have followed since then Tamil Report on the Agricultural Labour Enquiry, 1956-57 (Labour Bureau, 1960-62)

## Rural Labour Enquiry (NSSO)

NSSO have also conducted what is called Rural Labour Enquiry in its 18th round of Survey during 1963-64. The purpose was to collect data on employment, unemployment, under employment, wages, consumption, expenditure and indebtedness of agricultural and non-agricultural labour in the rural households. Data regarding rural income and expenditure are given in the NSSO Report.

## TRADE UNIONS

Labour Bureau collects returns from individual trade unions under Indian Trade Unions Act, 1926. The statistics relate to number of trade unions, their membership by state and by industry, and their income and expenditure. Similar information is given for employer's unions also.

These data consolidated and published in the Annual Trade Unions in India (Labour Bureau. 1946-47) and also in Indian Labour Statistics. The statistics cover only registered unions and even that not fully.

## Miscellaneous Labour Statistics

### Industrial disputes

Statistics of industrial disputes arising between employers and employees are collected by the state Labour Departments and the Regional Labour Commissioners on voluntary basis.

The statistics relate to 1) number of disputes classified by state, by industry and by case : 2) number of workers involved 3) man days lost and 4) the percentage distribution of disputes by duration.

The information is available in India Labour Statistics (Labour Bureau) and also in its Labour Year Book. These data do not include political, sympathetic and similar strikes.

## Absenteeism

The Labour Bureau collects data on absenteeism on the basis of returns. Filing of returns is voluntary. Therefore, the information that is obtained pertain to a selected larger units in the category of mining, manufacturing and plantations. The data are given in Indian Labour Statistics.

## Hours of Work

Hours of Work statistics are available, so far as the organised sector is concerned, in the Indian Labour Statistics. Annual data of total Man-hour worked in each industry will be found in the Annual Survey of Industries.

The scope of these date is limited to the organised factory sector where the units are required to submit returns under the provisions of the Factory Act 1948. Some data are available in certain reports of the NSSO, regarding hours of work for the unorganised sector.

## Industrial Injuries

Statistics for industrial injuries are available for factory, mining railways and dock workers by category as fatal, non-fatal and serious with industry breakup. These statistics are available in

1) Indian Labour Statistics (Labour Bureau, 1959 annual)
2) Annual Reports of the Railway Board.
3) Report on the working of Indian Dock Labourers Act, 1934.
4) Annual Reports of the Chief Inspector of Mines.

## Gaps in Labour Statistics

The user of statistics should note their deficiencies. These are :

i) Employment data are poor and are not comparable over time and regions. The scope and coverage of data have always changed from time to time. The coverage needs further extension. Employment statistics are available regularly for factories, mines and government undertakings; others lie outside the scope. The available data account for just 10 % of the total employment in the country.

Unemployment statistics are scanty; no regular series are available. Data are gathered from employment exchanges which work in urban areas only. The data so collected cannot be said to be complete.

ii) Wages and earning statistics are also not adequate. Here also the coverage of the available data is limited to the organised sector and a part of the unorganised sector. In many cases, statistics are inadequate in respect of 'other emoluments'.

iii) The time lag in publication of data, as usual is also unduly long, and the utility thereof is marred.

## POPULATION STATISTICS

When we talk of population statistics, we mean thereby factual data on the number of persons living in the country, division into male and females, material status, working and non-working population, nationality, religion, age, etc.,

There are three methods which are generally employed to collect population statistics.

1) **Population Census :** In most countries a population census (complete count of all the people alive on a particular day) is conducted once in ten years under state direction.

2) **Registration of Births and Deaths :** The recording of vital events such as births, deaths and marriages helps us to estimate the probable strength of the population on a particular data.

3) **Adhoc Survey :** Adhoc surveys known as demographic surveys are conducted by official agencies of a particular region to collect the required information.

The population of a country can be enumerated accurately with help of the vital statistics alone, provided a correct and up-to-date record of births and deaths is maintained in the country. However, generally, these records are not complete and therefore, vital statistics alone fail to provide an accurate idea about population growth. Population census deals with actual counting of all persons located in each area of the country on a particular day and collecting details regarding their age, sex, economic standard, occupation, nationality, etc. Population census, provides us with the number of individuals inhabiting a country on a particular date while record of vital statistics provides us with such information and at a time we required. Though population census and vital statistics are two important methods for estimating the population growth, generally both the methods are used together. The demographic survey not an independent method and generally it refers to a particular region or sector only. A brief description of population census and vital statistics is given below.

### Population Census

According to the U.N. document 'Principles and Recommendations for National Population Censuses' a census of population may be defined as "the total process of collecting compiling and publishing demographical economic and social data pertaining at a specified time or times to all persons in a country or delimited territory".

Census today is not merely counting of heads. Census data are extremely useful, may be indispensable, in framing suitable policies and in the absence of this basic data most of the policies would be a leap in the dark and planning almost failure. Such statistics are useful for economists, sociologists, businessmen, trade, industry and commerce. They provide valuable material for public health authorities for disease control and rationing authorities for estimating the requirements of food supply.

The economist by studying the trend of population growth, its occupational structure and increase in rural and urban population is in a position to make valuable suggestions to the administration of the country. He can trace the connection between population growth and food supplies, occupational changes and grant of protection of industries, increase in urban population and decay in rural population and make suggestions for changes in the state policy for proper and balanced growth of the country.

The sociologist studies the directions of reforms in matters of disease control, material status, age at which people should marry, family planning etc. Schemes of social insurance, urban and rural welfare, all dependant facts returned in population statistics. The knowledge about consumers, their classification according to income, location, etc., helps businessmen in launching sales campaign, market expansion possibilities, organising publicity drives and deciding the media of advertisement. Population statistics also serve trade, industry and commerce of the country to a considerable extent. Hence, it is rightly said that population census is not merely counting of heads-it supplies a heap of valuable information which is of great significance in different spheres.

## Methods of Conducting Population Census

Broadly speaking, there are two methods of conducting a population census.

1. De facto method, and
2. De jure method

In the de facto method all persons are counted in the area, where they are physically found on the day or night of the census. Thus under this method, all persons living in the country are counted simultaneously wherever they are found on the census night. In the de jure method, on the other hand, the population of each area is defined as persons who usually reside in the area regardless of their actual location at the census date. Thus, under this method, people who are temporarily away from their normal place of residence on the census night are enumerated at the place of their normal residence. It should be noted that the de facto method is also known a one night enumeration method" Or, simply "data system" for the reason that in this method enumeration has to be done simultaneously in one night. On the other hand, the de jure method is also referred to as "period system of enumeration" for the reason that the census enumeration is not done in a single night but is spread over a period though figures relate to a particular date.

Some countries follow de facto method, others de jure method while still in others various combinations and modification of these concepts are used. The merits and demerits of these methods are discussed below.

## Merits of de Facto Method

i) This method of enumeration is simple. Enumerators or respondents can simply be instructed to report all persons who are present in each dwelling or other place of enumeration at the specified time. No elaborate instructions are required with reference to the inclusion or exclusion of persons who are away from home under different circumstances or for different reasons.

(ii) The other advantage of the de facto method is the objectivity of definition. There are no borderline cases of persons who may be enumerated or not depending upon the judgement of the enumerator or respondent.

(iii) "For purposes of international comparability, it is desirable that a de facto enumeration may be made, that is, count on all persons present in the country at the time of enumeration. Any data on a de jure basis which may be desired should be detained in addition to the de facto data" - Recommendation of the International Agency.

(iv) The de facto concept is especially, advantageous from the standpoint of international comparability, for it is an unequivocal standard that can be applied universally without regard to difference in local conditions.

(v) De facto figures would be useful to municipal authorities analyzing the requirements for policy protection, medical services hygienic and transportation and to businessmen estimating the demand for products or service purchased by residents and visitors alike.

## Limitations of De Facto Method

i) The simplicity of the concept is marred by the exceptions for military and diplomatic personnel which have been recommended by international agencies. Under the de facto method we also count those persons who do not actually belong to this country but are linking in this country and we exclude those who are living out side the country. Thus we do not get the actual figure, we include a portion of those people in whom we are not interested and exclude a portion of those people in whom we are interested.

ii) Census cannot be completed just overnight. A large number of persons will have to be employed which will mean great amount of expense. If the census-taking is spread over a number of days it would end in duplication, etc., Thus de facto population can also be defective.

To remove this defect what is done generally is that on the first ten or fifteen days or any other number of days (varying in different countries according to their practice) population is recorded and the corrections are done on the census day i.e., recorded details about the population are taken on a particular day to the various houses for corrections, if any.

iii) Again, there is a difficulty with regard to the persons moving in trains, etc., from one place to another-how they should be counted and what place of their residence should be taken, whether that place where the train is or that place where they are going or otherwise the place from where they are coming.

## Merits of De Jure method

i) The de jure population is the permanent population, which is what is desired.

ii) A de jure enumeration may give more accurate results than a de facto one, if the enumeration is to be spread over a considerable time interval. Among the mobile population, the dangers of double counting or omission multiply as the period of enumeration is extended. In particular

a de facto count of a very mobile population is subject to serious error is that the enumeration is spread over a long period. A de jure count under conditions is also likely to be inaccurate; but the errors in this case may be smaller if most of the people who move during the interval to a stable place of usual residence.

iii) A de jure enumeration is necessary in some countries to fulfill certain legal requirements notably to provide statistical bases for the appointments of electoral representatives, tax assessments, monetary grants, etc., among the area of the country. In some countries the legal authority for the census explicit require a de jure enumeration.

iv) Local requirements for population statistics in connection with the planning and administration of certain social services such as housing and education programmes which apply only to the resident population are better fulfilled, by the de jure than by de facto figures. There are also other uses for which de jure figures seem preferable, including perhaps the population of base figures for the calculation of birth and death rates.

## Limitations of De Jure Method

i) In a de Jure enumeration more complicated definition and instructions are necessary to ensure complete and consistent reporting. Generally speaking, the purpose in this kind of enumeration is to allocate each individual to his place of residence, though there are many difficulties in the way. The main difficulty is that persons do not have a fixed single place of abode.

ii) There are greater changes of errors and omissions and duplication. Examples are children away at school, individual families that have left home for long or indefinite periods to travel or work elsewhere, people who have no fixed residence, families that maintain two residences, families that maintain two residences one in the city and another in the country.

Even though detailed instructions and other special measures are taken to ensure a uniform basis for their enumeration, some persons are likely to be omitted or counted more than once. Even those persons who have a definite place or residence, but who are absent at the time of the census may be omitted, counted more than once or assigned to the wrong area, if the persons who report for them are not well informed.

For a country as a whole, the difference between a 'de facto' or a 'de jure' total is to be small under ordinary circumstances. The difference between the number of residents temporarily outside the country and the number of non- residents temporarily within it ordinarily amounts to only a small percentage of the total population. Under same conditions, however, the difference may be substantial.

De facto definition is much more suitable then a de jure for uniform application in different countries. The purpose for which de jure statistics are more appropriate, appears to be primarily rational and internal rather than international.

A better method of obtaining both the de facto and de jure count of persons temporarily out of the country, is to enumerate at each dwelling, both the persons who usually reside there but are

temporarily away and those who are present but have a usual place of residence elsewhere. These two categories must be separately identified on the census schedule. The accuracy of the results may be improved by recording for the former group, their location at the time of the census and for the latter, their usual place of residence.

## Methods of Enumeration

There are two important method of enumeration

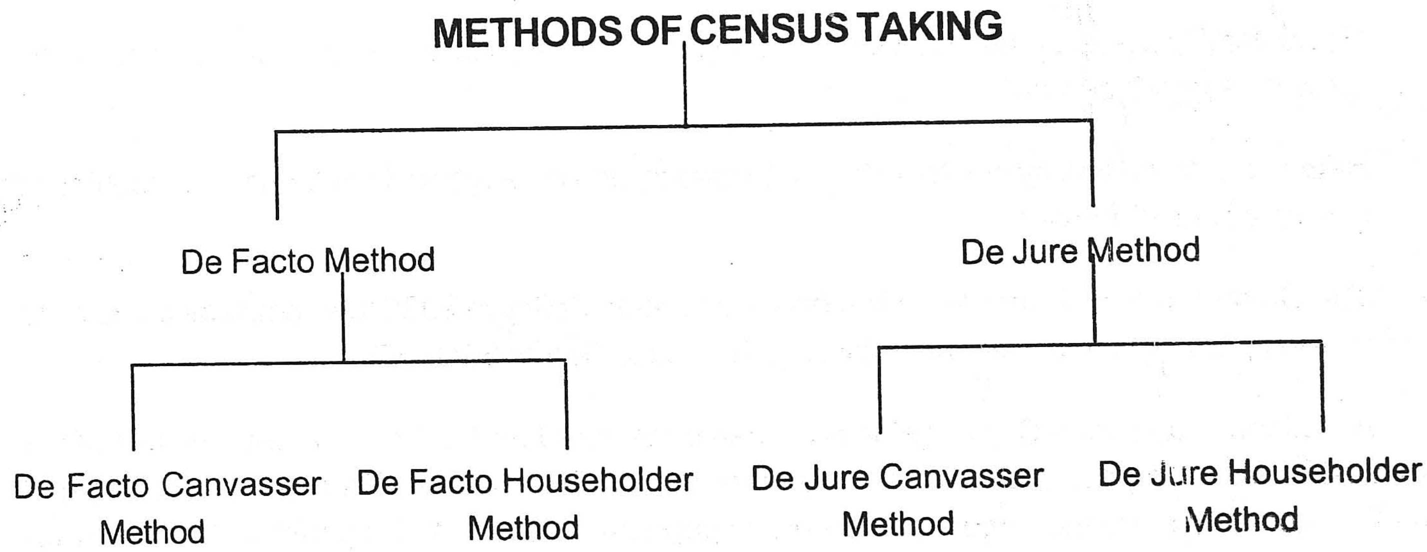
1. CANVASSER METHOD or enumeration by appointed enumerators. The information is obtained by a personal interview and entered on the schedule by officially designate enumerators.

2. THE HOUSEHOLDER METHOD OR SELF-ENUMERATION

Under this method schedules are distributed to all households, along with the instructions for filling them. The head of the household is responsible for entering the information and the schedules are collected by officially designated persons.

A combination of the two methods has also been used, part of the information being filled in by respondents, and part by the officially designated enumerators.

The following chart gives details of the methods of census taking.




## Population Census of India

The Indian Census today is universally acknowledged as the most authentic and comprehensive sources of information about the land and people. The history of census-taking in India dates back to the year 1872 when the first census was taken. This census was Limited in coverage of geographical territory and information collected was also not complete. The first complete census of the whole country was taken on February 1, 1881 and since then a census of population takes place every 10 years. The last census relates to the year over the last 110 years. The census operations have been keeping up with the varied demands of changing times. A greater degree of sophistication has come into our method of collection; tabulation and analysis of statistics data during this period. The last census was taken up in the year 2001.

In developing countries like India, the periodical population census is the best source for obtaining a frame of area units suitable, for carrying out sample surveys covering various fields such as demography, agriculture, social and economic statistics during the inter-census period. In fact; by providing a suitable sampling frame a population census facilitates through sample surveys, the collection of a host of statistics at national and regional levels, all of which cannot be furnished by the census itself except for some basic items.

## Population Problem - An Evaluation

During 1981-91, the population of India grew from 683 million in 1981 to 844 million in 1991 indicating an increase of 161 million during the period. The rate of growth of population had slightly declined to 2.11 percent during 1981-91. During the next 8 year period population grew from 844 million to 987 million an increase of 120 million. By the turn of the 20th century Indian population has crossed 1000 million mark. The high growth rate of population can be explained interns of a persistently high birth rate but relatively fast declining death rate.

The Ninth Five Year Plan has identified three principal factors contributing to the growth of the population.

They are

1) Large size of the population in the reproductive age group being 144.2 million in 1997.

2) Higher fertility due to unmet need for contraceptives and its estimated contribution to population growth being 20 percent.

3) Higher unwanted fertility due to high infant mortality and its estimated contribution to population growth being 20 percent.

The Government has adopted a National Population Policy in 2000 to encourage a two child norm for every family with an aim at stabilising the population at 2046 AD.

A National Population Commission has been constituted by the Government. The aim of the commission is to bring about a decline in the birth rate in Bihar, Madhya Pradesh, Rajasthan and Uttar Pradesh which are the major contributors to population growth in the country. The outlay for the Ninth Five Year Plan has been increased to Rs. 15,120 crores for the Department of Family Welfare.

## Importance and Utility of Population Census

As part of integrated programme of data collection, the population census is useful for administrative purposes and for many aspects of economic and social research and planning. The following is an account on its importance and utility:

### 1. In Administration and in Policy Decision

One of the most basic administrative uses of census data is the demarcation of constituencies and the allocation of representation on governing bodies like Parliament and Assemblies

Consideration of question of employment and manpower programmes, migration, housing, education, public health and welfare, social services, economic and social planning and numerous other aspects of the life of the country can be easily decided with the help of the data provided by the census.

## 2. In Doing Research

The population census provides indispensable data on the changing patterns of urban - rural concentration, the development of urbanised areas, the geographic distribution of population according to occupation and education; the sex and age structure, social and economic characteristics of population. These data are important both to research and practical problems of industrial and commercial growth and management.

## 3. In Business and Industry

Reliable estimates of consumer demand for variety of goods and services depend on accurate information on the size of the population and its distribution. These characteristics heavily influence the demand for various commodities and services. Since the local availability to labour for production and distribution of commodities is important in determining the location and organisation of enterprises this calls for the need of the census data.

## 4. In Planning

The census data is indispensable for social and economic planning. It is extremely useful for all types of planning like national income estimates, formulation of policies, health, agriculture and development of road and railways, housing needs, analysing the growth of consumer demand and savings etc., in the process of economic development .

## Drawbacks in the Indian population census

The following are some of the major drawbacks in our population census :

### 1. Imcomplete coverage

Our population census does not cover certain people like homeless persons and wanderers. This leads to an under estimation of population figures.

### 2. Untrained census staff

The census work is done by the officials and enumerators during their out-of office hours. They are given training only for a short period. Their experience is not sufficient to take up the responsibility of collecting information.

### 3. Indifferent attitude of people

Indian census is described as "a comet which appeared on the Indian horizon every ten years and after two or three years of activity passed away unnoticed". This damaging remark is due to people's indifference towards the census operations. It is quite essential to keep the public well informed as to what the census aims at and the usefulness of census to the nation.

## 4. Age Returns

Correct data on age is lacking in Indian population census. People's ignorance, indifference and illiteracy make the age returns incorrect. Accurate age returns are essential to calculate the mortality and other rates of population.

## 5. Classification of occupation

The absence of an exhaustive classification of occupation is another serious drawback in population census.

Inspite of the above defects the Indian census methodology has stood the test of time.

## II Vital Statistics

Vital statistics include statistics of birth and deaths. In India the Births, Deaths and Marriages Registration Act of 1886 provided for the voluntary registration of births and deaths. Some of our states make registration compulsory through their own acts. But these data are grossly inaccurate, misleading and unfit for statistical analysis because there is no uniformity in the procedure of registration in various states. In rural areas, the village headman and in urban areas the municipal authorities are to record the births and deaths. The collection of these data by these authorities is not dependable. However, recently the vital statistics collection has been taken over by the Registrar General and Census Commissioner. (These vital statistics are now provided in the Statistical Appendices to the Annual Report of the Director General of the Health Services)

## Shortcomings of Vital Statistics

The position with regard to the availability of vital statistics is far from satisfactory and efforts to improve them are to be steped up. The main limitations of these statistics are :

## Part - Time Agency

Most of the staff associated with the collection of vital statistics in India are on a part-time ex-officio basis. Part-time officers have been appointed from the Revenue, the Police, the Health and the Panchayat Departments on the basis of availability. Such employees find little time to attend to the duties connected with data collection and processing of data pertaining to vital statistics and as such the data are half-heartedly collected and cannot be much relied upon.

## 2. Delay in Publication

There is no whole-time agency upon whom reasonable standard of efficiency can be imposed and enforced and as such the publication of vital statistics is considerably delayed. Most local bodies who are primarily responsible for collecting vital data have prescribed time limits ranging from one day to 14 days within which the vital event must be reported. But failure to report this in time is not taken seriously and there is no legal provision for action against those who fail to report a vital event with the prescribed time.

### 3. Inaccuracy

Vital statistics in our country are not very accurate. It is because of two reasons: First, there is no legal sanction authorising of vital data; and secondly, the officials and the persons responsible for supplying the information feel immune from action.

### 4. Lack of uniformity

There is no uniformity in the procedure of registration in various states. It is necessary to collect the statistics in a well organised and uniform fashion throughout the country. In England the Ministry of Health is responsible for the collection of those figures.

### 5. Non-co-operation from the Masses

People do not take any interest and fail to realise the significance of vital statistics. In particular, this difficulty is badly experienced in rural areas. In urban areas this problem exists to a lesser degree.

## Suggestions for Improvement

A number of committees and commissions have suggested measures to bring about improvement in vital statistics such as the Central Advisory Board of Health (1939), the Health Survey and Development Committee (also known as the Bhor Committee of 1946). The Vital Statistics Committee (1949), the Committee Study and Report on patterns of statistical units for Health Departments (1960). Some of the suggestions are:

1. Compilation of vital statistics should be done on a uniform basis throughout India. The Central Government should pass the Indian Vital Statistics Act authorising the collection of data and making it obligatory upon individuals to supply the data.

2. Compilation of vital data and their processing should be done on a uniform basis throughout India. The Central Government should be centralised and mechanised and standard proforma should be used to this end.

3. Training should be given to the registration staff particularly at the field level in the compilation of data.

4. Public should be educated to realise the significance of vital data.

5. A certificate of registration must be issued to the information of the head of the family in which the vital event takes place.

6. The vital statistics should be closely linked with the population statistics.

In view of the great significance of family planning, it would be desirable to continue to collect regularly statistics of family planning efforts, such as number of family planning clinics, number of sterilisation operation performed, number of loop insertions, distribution of contraceptives etc. The data should be available not only by states and rural-urban areas but also size classes of household income and by other meaningful socio-economic classifications.

# DEMOGRAPHIC SURVEYS IN INDIA

## Introduction

Demographic surveys are meant to collect information about particular problems in certain regions and in many cases, they fill up the gaps of data relating to itner-census periods. In India we have both population census and vital statistics. Demographic surveys have been started only very recently.

## Survery of fertility of women in India

The pattern of fertility of women in India has been found to be a distinctive one according to a country-wide sample study of births and deaths conducted by the Registrar General of India in 1952-53 and 1953-54. Fertility in India follows a pattern which is entirely different from the pattern experienced in other countries of the world. In our country, the fertility is low in the age group 14-19 but it rises very sharply in the next age group of 2024 and rises slightly in the age group 25-29. After this there is a decline.

## SUGGESTED FURTHER READINGS


1. Asthana B.N. and Sri Vastava. S.S. Applied Statistics of India, Ed. Allahabad, Chaitanya Pub. House 1972.
2. Bansil, Puran Chand. Agricultural Statistics in India. ed. 2 New Delhi, Arnold-Heinemann. 1974.
3. Gupta, D.B. and Premi, M.K., Sources and Nature of the Official Statistics of the Indian Union. Delhi, Ranjit Printers. 1960.
4. Gupta.S.P. Statistical Methods, Sultan Chand and Sons New Delhi.
5. Rao.C.R., ed. Data Base of Indian Economy: V.I. Proceedings of a seminar organised by the Indian Econometric Society, New Delhi Statistical Publishing Society, Calcutta, 1972.
6. Saluja, M.R. Indian Official Statistical System, Statistical Publishing Society Calcutta, 1972.
7. Shukla, M.C. and Gulshan, S.S. Statistics Theory and Practice, S. Chand & Company Ltd., New Delhi. 1986.


## Model Questions

1. Explain the statistical availability of data on labour and trade unions.
2. Describe the availability of population statistics in India.
3. Write an appraisal on the population census in India.